

Open Research Online

The Open University's repository of research publications and other research outputs

Statistical analysis of quantitative seroepidemiological data

Thesis

How to cite:

Kafatos, George (2011). Statistical analysis of quantitative seroepidemiological data. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 2011 The Author



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.0000ee3e>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

George Kafatos, BSc (Hons), MSc

**Statistical analysis of quantitative
seroepidemiological data**

**Thesis submitted for the award of
Doctor of Philosophy in the Mathematics and Statistics discipline
of the Open University**

May 2011

**The Open University
Health Protection Agency**

Abstract

The material for this thesis is based on the European Sero-Epidemiology Network 2 (ESEN2), a study funded by the European Commission. As part of the ESEN2 project, age-specific population seroprevalence of 8 antigens tested in 22 national laboratories was estimated and compared between countries. To achieve harmonised serological results, a reference panel was tested by each participant laboratory. Each laboratory's panel results were regressed against the reference centre, thus obtaining standardisation equations. These equations were used to convert the quantitative measurements of the serosurveys into common units that were subsequently classified into negative (susceptible) or positive (protected) according to a serological cut-off.

The aim of this thesis was to further develop and validate the methodology for standardising serological outcomes, and to propose alternative methods for achieving comparable population seroprevalence.

As part of this thesis, a statistical algorithm was established to standardise serological results. Censored regression methods were considered to account for measurements outside the assay detection range. The impact of standardisation on seroprevalence was examined.

Mixture modelling of the serological results was proposed as an alternative method to standardisation for estimating seroprevalence. Although mixture modelling may provide better seroprevalence estimates in certain situations, it is heavily dependent on model assumptions, mainly of well-separated underlying distributions.

In terms of seroprevalence estimation using standardisation, the validity of the assay cut-off point was examined. A method for re-estimating cut-offs was proposed based on mixture modelling that improved seroprevalence estimates under certain distributional assumptions.

The impact of variability occurring due to serum testing in batches (plate-to-plate variability) on seroprevalence was assessed. The method currently used by the laboratories was examined, and a new method was proposed to adjust for this based on mixture models.

In conclusion, the standardisation method used for the ESEN2 project was validated and some improvements were proposed.

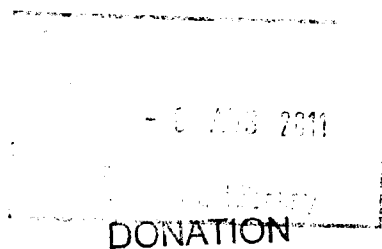


Table of contents

Abstract..... 2

Table of contents 4

List of Figures 10

List of abbreviations 17

Acknowledgements 19

Chapter 1: An introduction to seroepidemiology 20

 1.1 The burden of infections..... 21

 1.2 Vaccine-prevention..... 23

 1.3 Infectious disease surveillance 24

 1.4 WHO targets for vaccine-preventable diseases in the European region..... 25

 1.5 Epidemiology using serological data..... 30

 1.6 The European Sero-Epidemiology Network (ESEN) project..... 32

 1.7 Limitations of serosurveys..... 33

 1.8 The method of standardising serological results..... 34

 1.9 The European Sero-Epidemiology Network 2 (ESEN2) project..... 34

 1.10 Other international seroepidemiology studies 35

Chapter 2: Serological assays, vaccines and seroprevalence 37

 2.1 The immune system..... 38

 2.2 Diagnostic tests..... 40

 2.3 Vaccines..... 47

 2.4 Prevalence and incidence definitions..... 48

 2.5 Estimating population prevalence using serosurveys 50

 2.6 Estimating seroprevalence for a population under an existing vaccination programme 51

2.7 Correlates of protection.....	53
2.8 Seroprevalence is not always equivalent to proportion protected.....	55
Chapter 3: Aims and objectives	57
3.1 PhD project rationale	58
3.2 Overall aim.....	58
3.3 Specific objectives	59
3.4 Statistical methods	59
3.5 Developing further the standardisation methodology	61
3.6 Impact of standardisation on seroprevalence	62
3.7 Mixture models	62
3.8 Assay cut-offs	62
3.9 Plate-to-plate variability.....	63
3.10 Statistical software	63
3.11 Data source.....	64
3.12 Publications arising from the thesis	64
Chapter 4: Standardisation of serological results	66
4.1 Comparability of national seroprevalence estimates.....	67
4.2 An example comparing seroprevalence	68
4.3 Panel construction.....	69
4.4 Comparison of first and second panel test results	70
4.5 Selecting standardisation equations	73
4.6 Outlier investigation.....	75
4.7 Selection of regression model.....	77
4.8 Effect of results outside the assay detection range.....	82
4.9 Standardisation algorithm	85

4.10 Back-standardisation.....	86
4.11 Standardised vs. non-standardised seroprevalence estimates	87
4.12 Conclusions	90
Chapter 5: Censored Serological Data.....	92
5.1 Assay detection limits.....	93
5.2 Types of censored data	94
5.3 Deletion and simple substitution methods	95
5.4 Simulations for regression equations.....	98
5.5 Comparison of deletion and simple substitution methods using simulations.....	100
5.6 Censored regression.....	106
5.7 Method of Multiple Imputation (MI).....	108
5.8 Confidence intervals for MI estimates.....	109
5.9 Example and simulations.....	111
5.10 Investigation of regression assumptions	117
5.11 Interval censored regression	126
5.12 Method comparison when censoring occurs on the independent variable.....	131
5.13 Conclusions	134
Chapter 6: Comparison between standardised and non-standardised serological results.....	137
6.1 Serosurvey construction and national sero-profiles.....	138
6.2 Classification into negative, positive and equivocal results	139
6.3 Obtaining standardised and non-standardised sero-profiles	141
6.4 Between-laboratory and within-laboratory variability	146
6.5 Impact of standardisation on sero-profiles.....	147
6.6 Examples of differences between standardised and non-standardised seroprevalence	153

6.7 Variability due to different assays	158
6.8 Variability arising due to different antigens.....	159
6.9 Variability due to different age groups	160
6.10 Failure of standardisation method.....	161
6.11 Multivariable analysis of seroprevalence differences	162
6.12 Conclusions.....	168
Chapter 7: Mixture models as an alternative method to standardisation	170
7.1 Introduction - A standardisation limitation	171
7.2 Background on mixture modelling	171
7.3 A simple mixture model.....	172
7.4 Model diagnostics	175
7.5 Example of a mixture model (naïve model).....	180
7.6 Example of model diagnostics (naïve model).....	181
7.7 Varying mean and standard deviation by age group.....	187
7.8 Shape of underlying distributions	189
7.9 Model diagnostics (final model)	193
7.10 Seroprevalence estimation using standardisation.....	194
7.11 Model comparisons	196
7.12 Functional relationship of mean and standard deviation with age	197
7.13 Functional relationship of skewness parameter with age	201
7.14 Seroprevalence comparison for 14 age groups	205
7.15 Conclusions.....	206
Chapter 8: Estimating serological cut-offs for sero-epidemiology.....	210
8.1 Introduction.....	211
8.2 Assay cut-offs as specified by the manufacturers	211

8.3 ROC curve analysis	212
8.4 Are assay cut-offs appropriate for sero-epidemiology?.....	218
8.5 Estimating assay cut-offs using mixture models	222
8.6 Estimating an age-specific cut-off for the Spanish VZV serosurvey	225
8.7 Application of the estimated cut-offs to other national serosurveys.....	229
8.8 Impact of cut-off estimates on population seroprevalence	233
8.9 Method comparison using simulations	236
8.10 Advantages and disadvantages of each method.....	245
8.11 Conclusions	246
Chapter 9: Seroprevalence estimation taking into account plate-to-plate variability....	248
9.1 Introduction	249
9.2 Rubella sero-survey	249
9.3 Optical Density (OD) measurements.....	252
9.4 Control samples	255
9.5 Calibration curves.....	258
9.6 Estimating plate-to-plate variability using the control samples.....	261
9.7 International Units (IU/ml).....	262
9.8 Mixture modelling using IU/ml	265
9.9 Comparison between existing methods	266
9.10 Mixture model on OD measurements adjusting for plate-to-plate variability	267
9.11 Conclusions	271
Chapter 10: Discussion	273
10.1 Introduction	274
10.2 Standardisation methodology	275
10.3 Seroprevalence estimation	278

10.4 Limitations of the standardisation methodology	279
10.5 Extending mixture modelling methodology: current work	280
10.6 Future applications	282
10.7 General conclusions	283
Reference List	285
Appendix I Programmes.....	305
(A) Programmes for censored data	306
(B) Programme for multiple imputations.....	310
(C) Programme for mixture models	314
(D) Programme for cut-off estimation using R (Equation 8.5)	316
Appendix II Publications arising from the PhD	320

List of Figures

Figure 1.1 A map of Europe highlighting the ESEN2 participant countries 35

Figure 2.1 Innate and adaptive immune systems 39

Figure 2.2 Test tubes in a centrifuge machine 41

Figure 2.3 Scientist uses a multi-channel Pasteur’s pipette..... 41

Figure 2.4 An ELISA plate with 96 wells 43

Figure 2.5 Spectrophotometer for measuring the intensity of colour for an ELISA test 44

Figure 2.6 Relative rubella IgG and IgM levels after onset of rubella symptoms 49

Figure 2.7 Titre distributions for infected and non-infected individuals and positive /
negative cut-off 51

Figure 2.8 Titre distributions for infected, vaccinated and non-infected / non-vaccinated
individuals..... 53

Figure 4.1 Age-specific seroprevalence estimates of VZV antibody for the Netherlands and
Italy (non-standardised)..... 69

Figure 4.2 Comparison between first and second panel tests 72

Figure 4.3 Outlier investigation for Lithuanian measles results regressed against the
Germany reference centre 77

Figure 4.4 Comparison between linear and quadratic curves for the standardisation panel
tested against measles by the Slovakian and the German laboratories 79

Figure 4.5 Comparison of different regression lines for the HAV panel tested by the Czech
Republic and Greece laboratories 81

Figure 4.6 Influence of censored results on the regression line - VZV panel results for
Belgium plotted against the Spanish reference centre’s results..... 84

Figure 4.7 Influence of censored results on the regression line - HBV panel results for
Slovakia plotted against the Greece reference centre’s results 85

Figure 4.8 Flow-chart summarising the standardisation algorithm.....	86
Figure 4.9 Standardisation curve for Lithuania measles panel results	88
Figure 4.10 Seroprevalence estimates for Lithuania measles results	89
Figure 4.11 Age-specific seroprevalence estimates of VZV antibody for the Netherlands and Italy (standardised).....	90
Figure 5.1 Regression curves using deletion and simple substitution.....	97
Figure 5.2 Comparison between simple substitution and deletion methods after 1,000 simulations generated from $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$	102
Figure 5.3 Comparison cut-off estimate between simple substitution and deletion methods for different proportions of data censored generated from the model $y = -0.4 + 1.1x +$ $0.12x^2$ with $\sigma = 0.26$	104
Figure 5.4 Comparison between simple substitution and deletion methods after 1,000 simulations generated from $y = -0.4 + 1.1x$ with $\sigma = 0.26$	105
Figure 5.5 Comparison of cut-off estimate between simple substitution and deletion methods for different proportions of data censored generated from the model $y = -0.4 +$ $1.1x$ with $\sigma = 0.26$	106
Figure 5.6 Regression curves for Israel HAV panel using simple substitution, censored regression and MI	113
Figure 5.7 Comparison between censored regression and MI methods from 1,000 simulations generated from $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$	114
Figure 5.8 Comparison of cut-off estimate between the censored regression and MI methods for different proportions of data censored generated from the model $y = -0.4 +$ $1.1x + 0.12x^2$ with $\sigma = 0.26$	115
Figure 5.9 Comparison of cut-off estimate between the censored regression and MI methods for different proportions of data censored generated from the model $y = -0.4 +$ $1.1x + 0.12x^2$ with $\sigma = 0.26$ using a cut-off of -1	116

Figure 5.10 Comparison between simple substitution and censored regression methods from 1,000 simulations generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$ when the normality assumption does not hold ($\lambda = 2$).....	118
Figure 5.11 Comparison of cut-off estimate between censored regression and MI methods for different proportions of data censored generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$ when normality assumption does not hold (varying λ)	119
Figure 5.12 Comparison between simple substitution and censored regression methods from 1,000 simulations generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.52$ and $\sigma = 0.26$ for measurements below and above $\log_{10}(0.01)$, respectively	121
Figure 5.13 Comparison of cut-off estimates between the censored regression and MI methods for different proportions of data censored generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.52$ and $\sigma = 0.26$ for measurements below and above $\log_{10}(0.01)$, respectively	122
Figure 5.14 Non-linear regression curves using deletion, censored regression and MI....	124
Figure 5.15 Comparison between censored regression and the deletion methods after 1,000 simulations generated from $y = -0.3 + 2.5 / [1 + \exp(-(-4) + 3.5x)]$ with $\sigma = 0.16$	125
Figure 5.16 Comparison between deletion, simple substitution and interval censored regression methods	128
Figure 5.17 Comparison between deletion, simple substitution and interval censored regression methods after 1,000 simulations from $y = -0.2 + x$ with $\sigma = 0.3$	130
Figure 5.18 Regression curves using deletion and 90% PIs for a panel comparison with x-censoring.....	133
Figure 6.1 Frequency distribution of VZV Spanish results by age group	140
Figure 6.2 Standardisation curve for Israel VZV results.....	142
Figure 6.3 Distribution of Israel's VZV serosurvey results by age groups	143
Figure 6.4 Sero-profile for Israel's VZV results before and after standardisation	144

Figure 6.5 Non-standardised results for Israel VZV serosurvey by age group classified into negative, equivocal and positive.....	146
Figure 6.6 Standardisation curve for Italian HAV results	154
Figure 6.7 Differences between standardised and non-standardised seroprevalence of Italian HAV results by age	155
Figure 6.8 Standardisation curve for UK rubella results	156
Figure 6.9 Differences between standardised and non-standardised seroprevalence of UK rubella results by age	157
Figure 6.10 Differences between standardised and non-standardised seroprevalence of Belgian measles results by age	158
Figure 6.11 Distribution of the mean seroprevalence differences.....	163
Figure 6.12 Distribution of the \log_{10} -transformed mean absolute differences.....	163
Figure 6.13 Multivariable model results of mean absolute seroprevalence differences by antigen and whether back-standardisation was used (adjusted for laboratories and whether labs used the reference assay)	167
Figure 7.1 Fitted model assuming constant mean and standard deviation for Belgian serosurvey tested against VZV	181
Figure 7.2 Pearson residuals by age group for the naïve model (excl. censored data).....	183
Figure 7.3 Fitted model and observed data with 95% bootstrap intervals for the naïve model.....	186
Figure 7.4 ECDF plotted against mixture distribution CDF by age group	187
Figure 7.5 Fitted model varying the mean and standard deviation of the mixture model positive component by age group	189
Figure 7.6 Fitted model with a skew normal positive component varying its parameters by age group	192
Figure 7.7 Pearson residuals by age group for the final model.....	193

Figure 7.8 Assay cut-off point and frequency distribution of VZV Belgium results by age group	195
Figure 7.9 Comparison between standardised and non-standardised VZV seroprevalence estimates by age	196
Figure 7.10 Distribution of mean and standard deviation of log ₁₀ IU/ml results by age confined to samples classified as “positive” by standardisation.....	199
Figure 7.11 Fitted mean and standard deviation of log ₁₀ IU/ml results by age for the skew normal component of the mixture distribution (past infections).....	203
Figure 7.12 Pearson residuals vs. log ₁₀ IU/ml by age group	205
Figure 7.13 Seroprevalence comparisons	206
Figure 7.14 Distribution of the standardised mumps serosurvey results for UK.....	208
Figure 8.1 Standardisation curve for data from Slovakia and the Spanish reference centre	213
Figure 8.2 Area under the curve for Slovakia VZV panel.....	215
Figure 8.3 Seroprevalence estimates for Slovakia using local assay results, standardisation and ROC curve analysis.....	216
Figure 8.4 Seroprevalence estimates using ROC analysis for Lithuanian measles results	217
Figure 8.5 Distribution of non-standardised serological VZV results for Slovakia and positive / negative cut-off point as defined by the assay manufacturer	219
Figure 8.6 Distribution of serological VZV results for Spain and positive / negative cut-off point as defined by the assay manufacturer	220
Figure 8.7 Distribution of standardised serological VZV results for Slovakia and positive / negative cut-off point as defined by the reference centre’s assay manufacturer	221
Figure 8.8 Standardised vs non-standardised seroprevalence estimates for Slovakia serosurvey	222
Figure 8.9 Pearson residuals for Spanish VZV serosurvey plotted against log ₁₀ IU/ml by age group.....	227

Figure 8.10 Comparison between the fixed cut-off given by the assay manufacturer and the age-specific cut-off as estimated using the mixture model estimates.....	228
Figure 8.11 Seroprevalence comparisons using different cut-off estimates	229
Figure 8.12 Flow-chart showing the process of estimating a cut-off point	231
Figure 8.13 Comparison between the cut-offs as estimated by the different methods.....	232
Figure 8.14 Observed data and fitted mixture model for Slovakia standardised serosurvey results	234
Figure 8.15 Pearson residuals for Slovakian VZV serosurvey plotted against \log_{10} IU/ml by age group.....	235
Figure 8.16 Comparison between cut-off estimates using the mixture models defined in Table 8.4.....	238
Figure 8.17 Comparison between cut-off estimates using the mixture models defined in Table 8.6.....	241
Figure 8.18 Comparison between cut-off estimates using the mixture models defined in Table 8.14.....	244
Figure 9.1 Rubella vaccination programme for the UK	250
Figure 9.2 The design of the serological plates used for the rubella tests by the UK laboratory	252
Figure 9.3 Frequency distribution of \log_{10} OD measurements for UK rubella serosurvey	253
Figure 9.4 Observed data and fitted mixture model distribution of \log_{10} OD measurements	254
Figure 9.5 Pearson residuals by age group for \log_{10} OD measurements	255
Figure 9.6 Observed data and fitted mixture model distribution of \log_{10} OD _A measurements	257
Figure 9.7 Pearson residuals by age group for \log_{10} OD _A measurements	257
Figure 9.8 Generated calibration curves for the 45 plates	260

Figure 9.9 Comparison between IU/ml given by the assay reader (\log_{10} IU/ml) and as estimated using the generated calibration curves (\log_{10} new IU/ml) 261

Figure 9.10 Comparison between \log_{10} OD and \log_{10} IU/ml by age group 263

Figure 9.11 Comparison between \log_{10} OD and \log_{10} IU/ml by age group after excluding plate 38..... 264

Figure 9.12 Observed data and fitted mixture model for \log_{10} IU/ml..... 265

Figure 9.13 Pearson residuals by age group for \log_{10} IU/ml 266

Figure 9.14 Observed frequencies and fitted mixture model by age and plate groups 269

Figure 9.15 Pearson residuals by age group and plate groups for \log_{10} OD..... 270

Figure 10.1 Distribution of the mumps serosurvey results for UK and assay cut-off..... 280

Figure 10.2 Distribution of the mumps serosurvey results for UK and Romania and reference centre's assay cut-off..... 281

Figure 10.3 Estimates of negative, vaccinated and past infected groups as well as MMR vaccine coverage by age group 282

List of abbreviations

AIC: Akaike Information Criterion

AUC: Area Under the Curve

anti-HBc: Hepatitis B core antigen

anti-HBs: Hepatitis B surface antigen

AUC: Area Under the Curve

BIC: Bayesian Information Criterion

CDC: Centers for Disease Control and Prevention

CDF: Cumulative Distribution Function

Cfi: Center for Infections

CI: Confidence Interval

CoP: Correlates of Protection

COR: Cut-Off Ratio

CRS: Congenital Rubella Syndrome

DALY: Disability-Adjusted Life Year

DL: Detection Limits

D_L: Lower Detection Limit

D_U: Upper Detection Limit

ECDF: Empirical Cumulative Distribution Function

ELISA: Enzyme-Linked ImmunoSorbent Assays

EP: Extreme sero-Prevalence difference

ESEN: European Sero-Epidemiology Network

ESEN2: European Sero-Epidemiology Network 2

HAV: Hepatitis A Virus

HBsAg: Hepatitis B surface Antigen

HBV: Hepatitis B Virus

HCV: Hepatitis C Virus

HIB: Haemophilus Influenzae type B

HPA : Health Protection Agency

HPV: Human Papillomavirus

IU/ml: International Units per millilitre

IQR: InterQuartile Range

MCMC: Markov Chain Monte Carlo

ML: Maximum Likelihood

MMR: Combined Measles, Mumps, Rubella (vaccine)

MP: Mean sero-Prevalence difference

MR: Combined Measles, Rubella (vaccine)

MSE: Mean Square Error

NT: Neutralisation Test

PHLS: Public Health Laboratory Service

OD: Optical Density

OLS: Ordinary Least Squares

PDF: Probability Density Function

PI: Prediction Interval

ROC: Receiver Operating Characteristic

VZV: Varicella Zoster Virus

VE: Vaccine Efficacy

WHO: World Health Organization

Acknowledgements

First of all, I would like to thank my supervisors for their excellent guidance. In particular, I would like to thank Paddy Farrington for his ideas, wisdom and some truly helpful and detailed comments and Kevin McConway for providing a pure statistical point of view not influenced by past knowledge of seroepidemiology. Last but not least, I would like to thank Nick Andrews for his invaluable guidance and expert advice, and also, for his continuous support and motivation throughout the project.

This thesis would not have been possible without the permission of the members of the ESEN2 group to use the serological data collected as part of the ESEN2 project.

I would also like to thank Nigel Gay for sharing his knowledge of mixture models, Irene Kafatos and Jordana Peake for proof reading the thesis, and Faiza Hansraj for looking at the sections referring to the immune system.

Finally, I would like to thank the Health Protection Agency for sponsoring this PhD. In particular, I would like to thank Prof. Elizabeth Miller, since it was her idea to start the thesis in the first place, and Dr. André Charlett, the head of HPA Statistics Unit, for his general support and for arranging the HPA sponsorship.

Chapter 1: An introduction to seroepidemiology

1.1 The burden of infections

Despite predictions of their eradication following the introduction of vaccinations in the 20th century, communicable diseases continue to cause health, economic, and social problems worldwide (Walsh CT & Fischbach MA, 2009). In 2004, infectious and parasitic diseases were estimated to be responsible for 13.8 million deaths or 23.4% of the global mortality. In particular, six diseases – pneumonia, tuberculosis, diarrhoeal diseases, malaria, measles and HIV/AIDS - caused 19% of global mortality (World Health Organization, 2008d).

When comparing infectious disease mortality rates, there are big inequalities by wealth and age groups. During 2004, infectious diseases were responsible for 55.9% of total deaths in the developing countries of Africa in comparison to 4.9% of total deaths in Europe. Infectious diseases were estimated to be the main cause of death for 55.9% of all deaths in children under the age of 15, compared to 15.2 % of total deaths for adults of 15 years or older (World Health Organization, 2008d). Although in 2004 the number of deaths caused by communicable diseases remained high, the proportion of deaths caused by infectious diseases was lower. In 1993, infectious and parasitic diseases were totalling 16.4 million deaths worldwide i.e. 32.2% of the global mortality.

For most infectious diseases the number of deaths decreased between 1993 and 2004, with some exceptions, the most significant being HIV/AIDS (World Health Organization, 1995). The number of deaths in 1993 and in 2004 for some of the most common infectious diseases is given in Table 1.1. Note that 2004 has been used as the reference year throughout, because it was then that the latest “The global burden of disease” report was published by the World Health Organization (WHO) that year (World Health Organization, 2008d).

Table 1.1 Global mortality estimates for infectious and parasitic diseases by cause through the decade 1993 - 2004

Cause	1993		2004	
	Deaths (thousands)	Percentage of all deaths	Deaths (thousands)	Percentage of all deaths
Childhood-cluster diseases	1,678	3.3	847	1.4
Diphtheria	3.9	0.0	5	0.0
Measles	1,160	2.3	424	0.7
Pertussis	360	0.7	254	0.4
Poliomyelitis	5.5	0.0	1	0.0
Tetanus	149	0.3	163	0.3
Diarrhoeal infections	3,010	5.9	2,163	3.7
Hepatitis B*	933	1.8	105	0.2
HIV/AIDS	700	1.4	2,040	3.5
Lower respiratory infections	4,110	8.1	4,177	7.1
Malaria	2,000	3.9	889	1.5
Tuberculosis	2,709	5.3	1,464	2.5
<i>Total - Infectious and parasitic diseases</i>	16,445	32.2	13,778	23.4
Total - All causes	51,000	100.0	58,772	100.0

* does not include liver cancer and cirrhosis resulting from chronic hepatitis B infection

Source: World health Organization (World Health Organization, 2008d;World Health Organization, 1995)

The mortality rates are only part of the infectious disease burden, since hundreds of millions of people especially in developing countries are disabled every year. A quantitative measurement of the overall disease burden that incorporates morbidity, named *disability-adjusted life year (DALY)*, was originally invented and is being used by WHO. Measured in DALYs, 6.2% of the total disease and injury burden of the world in 2004 was due to lower respiratory infections, 4.8% due to diarrhoeal diseases, 3.8% due to HIV / AIDS and 2.2% attributed to each of tuberculosis and malaria diseases. There were also large regional differences, with no infectious diseases reported within the five diseases

with the highest DALYs for high income countries, whereas four out of five diseases (lower respiratory infections, diarrhoeal diseases, HIV / AIDS and malaria) had the highest DALYs for low income countries (World Health Organization, 2008d).

1.2 Vaccine-prevention

Much of the mortality reduction of infectious diseases can be attributed to the success of vaccination programmes. An example is the reduction from 1,160,000 to 424,000 deaths caused by measles between 1993 and 2004 (Table 1.1). However, although it is estimated that immunisation averts 2.5 million deaths every year, a further 2.5 million could also be prevented by applying more effective immunisation campaigns (World Health Organization, 2008d; World Health Organization, 2009).

The immunisation strategies can vary depending on the aims of the vaccination programme. These can be grouped into the following three categories:

(i) *Control or containment* is the reduction of disease incidence, prevalence, morbidity or mortality to an acceptable level, so it does not constitute a public health problem. When the aim of the vaccination programme is the containment of the disease, selective immunisation targeting specific subgroups of the population can be used. Given that such subgroups are easily identifiable, this type of immunisation campaign can be of low cost (Dowdle WR, 1999).

(ii) *Elimination* is when the endemic transmission of the infection is interrupted locally. Examples of this are the measles vaccine, where measles elimination has been achieved in Western Europe, and the poliomyelitis vaccine, where polio elimination has been achieved in the Western hemisphere (Dowdle WR, 1999; Nelson KE & Masters Williams C, 2007).

(iii) *Eradication* means that both the disease and the organism cease to exist in the population. In other words, there is no risk of infection or disease in the absence of immunisation. The only example of eradication is that of smallpox. A campaign for global eradication of the disease that started in 1956, ended in 1977; the last case was reported in Somalia. This is considered by many as the biggest public health achievement of the last century (Nelson KE & Masters Williams C, 2007).

If the aim of the vaccination is the elimination or eradication of the disease, a programme of mass immunisation targeting early childhood is required to prevent natural infections. Such a programme does not need to achieve 100% vaccination *coverage* to accomplish its aim (coverage is defined as a proportion of the population being vaccinated) (Nelson KE & Masters Williams C, 2007). If vaccine coverage reaches a certain critical level, then the population (or herd) immunity will be sufficient to shield even susceptible (unprotected to the infection) individuals (Anderson RM & May RM, 1991; Nelson KE & Masters Williams C, 2007).

To achieve elimination or eradication of a disease, selective immunisation can also be employed to supplement mass vaccination campaigns when a subgroup of the population is considered to be particularly at risk. For example, a combined measles-rubella vaccine was offered to all UK schoolchildren as part of a “catch-up” campaign in 1994 following a rubella outbreak, despite the fact that a routine immunisation programme with the combined measles, mumps and rubella (MMR) vaccine was in place since 1988 (Vyse AJ *et al.*, 2002; Vyse AJ *et al.*, 2006).

1.3 Infectious disease surveillance

Since the end of the 19th century, many Western countries established *surveillance* systems for reporting cases of certain infectious diseases. Although these were initially set up as

early notification systems for outbreaks, the information collected has also been essential for epidemiological studies (Giesecke J, 2002). The main benefit of surveillance from an epidemiological perspective is the estimation of *incidence* i.e. the number of individuals acquiring the disease over a time period divided by the total population at risk. Incidence estimates can be used to examine the burden of diseases, their trends and the impact of intervention policies. Information on the number of disease cases can be collected in various ways such as physicians' reports on patients' diagnoses, diagnostic reports from microbiological laboratories or hospital records (Giesecke J, 2002).

1.4 WHO targets for vaccine-preventable diseases in the European region

As mentioned earlier, in Europe the infectious disease epidemiology is different to the developing world's. Many European countries have long-established vaccination programmes that reach over 90% population coverage for infections such as diphtheria, tetanus and pertussis (World Health Organization, 2009). However, the WHO estimates that approximately 32,000 people are still dying every year from vaccine-preventable diseases in Europe (World Health Organization Regional Office for Europe, 2008).

Following the success of poliomyelitis elimination in Europe (Vilayleck MS, 2002), the WHO Regional Office for Europe set up targets for the elimination of other vaccine-preventable diseases such as measles and rubella. Back in 1998, the plan was to interrupt indigenous measles transmission by 2007, and reduce the incidence of congenital rubella syndrome (CRS) in the European region by 2010 (< 1 cases of CRS per 100,000 live births) (Centre for Disease Control and Prevention, 2004). In 2005, these targets were updated to include measles and rubella elimination by 2010 (Spika JS, 2006). Given that the 2010 deadline was not reached, it has now been postponed to 2015 (Steffens I *et al.*, 2010).

A brief description of the clinical symptoms and the epidemiology of eight vaccine-preventable diseases (measles, rubella, mumps, pertussis, diphtheria, varicella zoster virus, hepatitis A and hepatitis B) in the European region are given below. Such a description is necessary, since the analysis carried out for this thesis was based on data on these infections, collected as part of a European project.

Measles

Measles is characterised clinically by fever, cough, coryza (running nose), and conjunctivitis (red eyes), followed by body rash. Complications occur in 40% of the cases, ranging from pneumonia and other respiratory infections, to diarrhoea and eye disease which can often lead to blindness (Nelson KE & Masters Williams C, 2007).

The measles vaccine has been part of childhood vaccination programmes for at least 20 years in Europe, with an overall estimated coverage rate of 94% (Martin R *et al.*, 2009). Although mortality has been greatly reduced in Europe (12 reported deaths in the European Union countries in 2005), measles remains a public health concern with a high number of cases and outbreaks reported in several European countries in 2006 and 2007 (Muscat M *et al.*, 2009; Eurosurveillance editorial team, 2007).

Rubella

Rubella (otherwise known as *German measles*) is a generally mild disease, mainly characterised by body rash. However, complications might occur when acquired by adults and especially women during pregnancy (World Health Organization, 2008c). Given that congenital rubella infection can lead to miscarriage, fetal death or the birth of an infant with CRS (symptoms include hearing loss, eye abnormalities, congenital heart disease and mental retardation), the rubella vaccination programmes were initially oriented towards women of childbearing age (Nelson KE & Masters Williams C, 2007). With the

introduction of the MMR vaccine, the immunisation strategy was altered to transmission prevention via childhood vaccinations. However, unless there is very high vaccine coverage, childhood immunisation can paradoxically lead to an increase in susceptible adults, and therefore, an increase of CRS due to a decrease of virus transmission and an accumulation of susceptible adult females (Nardone A *et al.*, 2008).

Mumps

Mumps, or *infectious parotitis*, is a disease that affects the salivary glands. It is generally a childhood disease, however, it can also occur in adulthood where complications are more likely to occur (World Health Organization, 2008b).

Despite the relatively high coverage of MMR vaccine, mumps outbreaks have recently been recorded in a number of European countries (Schwarz NG *et al.*, 2010; Vyse A *et al.*, 2007). Some of these outbreaks are taking place in highly immunised populations (Brockhoff HJ *et al.*, 2010; Vandermeulen C *et al.*, 2009). Such outbreaks challenge the effectiveness of the vaccines offered (protective effect of vaccine as estimated by post-license observational studies) and the appropriateness of the diagnostic methods (Vandermeulen C *et al.*, 2009; Vyse AJ *et al.*, 2006).

Pertussis

Pertussis (otherwise known as *whooping cough*) is caused by the bacterium *Bordetella pertussis*, and its main characteristics are mild respiratory symptoms. It affects mainly children and infection during infancy can cause severe symptoms and death. In some European countries adult cases are an increasing problem (de Melker H *et al.*, 1997). In the past, pertussis was a very common disease, however, the introduction of routine vaccination programmes, as early as the 1950s and 1960s for some Western European countries, has succeeded in greatly reducing the incidence (World Health Organization,

2005). Moreover, since mid-1990s, different types of vaccines (acellular pertussis vaccines) have been introduced in some European countries such as Sweden (Hallander HO & Gustafsson L, 2009) and the UK (Andrews N *et al.*, 2010). Despite the high vaccine coverage, pertussis notifications have slowly been increasing in a number of European countries the last few decades. The reduction of pertussis transmission has not been fully established yet due to the variety of vaccination programmes (Hallander HO & Gustafsson L, 2009; Wearing HJ & Rohani P, 2009).

Diphtheria

Diphtheria is caused by the bacterium *Corynebacterium diphtheriae*, and affects mainly the upper respiratory system. Although most cases are mild or asymptomatic, a high fatality ratio (> 10%) has been reported in recent outbreaks (World Health Organization, 2006).

After the introduction of diphtheria vaccines from the 1940s the number of cases reported in Europe was reduced from more than 100 per 100,000 population in 1940s to just 623 in total in 1980 prompting the WHO to set up a target of disease elimination in the region by 2000 (Galazka AM *et al.*, 1995; WHO Regional Office for Europe, 1996). However, following an epidemic that occurred in the 1990's in the newly independent states of the former Soviet Union, this target was postponed indefinitely (Hardy IR *et al.*, 1996).

Reasons for the epidemic might include deterioration of socioeconomic conditions, a large group of susceptible adults, a decrease in vaccine coverage and a migration increase (Galazka A & Tomaszunas-Blaszczyk J, 1997). There seem to be large discrepancies in the proportion of the population susceptible to diphtheria between different European countries (Edmunds WJ *et al.*, 2000; Kolodkina V *et al.*, 2006).

Varicella Zoster Virus (VZV)

Varicella (or *chickenpox*) is a mild childhood disease but can lead to complications if acquired by adults, neonates (severe neonatal varicella) or pregnant women (congenital varicella syndrome). In addition to the initial infection, VZV remains dormant within the nervous system, and may later in life cause herpes zoster, which often has serious complications, particularly in the elderly and *immunocompromised* (having the immune response attenuated, usually as a result of disease, malnutrition, or immunosuppressive therapy) (Pinot de Moira A & Nardone A, 2005).

Many European countries have recently introduced VZV vaccination campaigns targeting specific risk groups, such as neonates, premature infants, pregnant women and immunocompromised. Others are currently considering introducing VZV vaccination as part of a routine immunisation programme in children (Pinot de Moira A & Nardone A, 2005). However, a number of modelling studies have shown that such an immunisation strategy might result in an increase in herpes zoster and have instead suggested targeting the high incidence elderly group (van Hoek AJ *et al.*, 2009). Although VZV has been part of routine childhood immunisation in the USA since 1995, the impact of this programme on herpes zoster incidence has not been assessed conclusively (Reynolds MA *et al.*, 2008).

Hepatitis A Virus (HAV)

Hepatitis A is an acute liver disease caused by HAV, and although it is asymptomatic in the great majority of the cases (especially in children) and has low mortality, it is an important source of morbidity in South and East Europe (Cianciara J, 2000; World Health Organization, 2000).

Recent epidemiological studies suggest a decrease in HAV infections due to improvements in socioeconomic and hygiene conditions, and in some cases as a result of effective

vaccination. The decline of infection particularly during childhood means that there may be higher susceptibility in adulthood in the future where the symptoms are more severe. The WHO position on hepatitis A vaccines emphasises that the endemicity status of the country should guide the vaccination policy (Anastassopoulou CG *et al.*, 2009). In general, Eastern and Southern European countries have higher incidence rates than Western or Northern countries, therefore more widespread immunisation programmes are needed in the former (Jacobsen KH & Koopman JS, 2004).

Hepatitis B Virus (HBV)

Hepatitis B is a disease that affects the liver and may cause both acute and chronic disease (World Health Organization, 2008a). Similarly to HAV, the epidemiology of HBV varies across Europe, with higher incidence rates occurring in Eastern and Southern Europe. Given the mortality burden from HBV in Europe (7,000 deaths a year) the WHO called all countries to introduce universal HBV vaccination. By the end of 2004, 43 out of 52 countries in the European region had universal vaccination programmes in place (Nardone A *et al.*, 2009).

1.5 Epidemiology using serological data

In order to assess the progress towards the targets set by the WHO, and to optimise the immunisation strategies for the above diseases in Europe, vaccination programmes need to be evaluated and improved to amend any perceived weaknesses. More specifically, susceptible cohorts need to be identified and potentially targeted with additional “catch-up” campaigns. Vaccination programmes may be optimised by comparing their effectiveness among countries implementing a variety of immunisation strategies.

As mentioned before, incidence estimation is one of the main aims of surveillance systems as defined by WHO. Other objectives include collecting information on clusters of disease,

vaccine-related adverse events, immunisation coverage and possible accumulation of susceptible individuals (WHO Regional Office for Europe, 2009). Incidence estimates can be combined with vaccine coverage to monitor the effectiveness of a vaccination programme, or to estimate *prevalence* i.e. the proportion of individuals diagnosed with the disease at a specific time. However, a weakness of surveillance systems based on clinical cases is that they cannot identify non-symptomatic cases which for some pathogens account for a high proportion of infections.

An alternative approach is to estimate *seroprevalence* i.e. the proportion of individuals tested positive for past infection based on serology (a more extensive definition and discussion is given in the next Chapter) (Mosby, 2009). The population seroprevalence can be estimated through the collection and testing of a sample set of serum samples (alternatively, other body fluids can be used such as saliva or urine) that are representative of the population. The process of collecting and testing samples from individuals at risk aiming to determine susceptibility to a disease is called *serosurveillance*, and the corresponding studies *serosurveys*. The conduct (design and analysis) of serological studies on individuals and populations that are used to monitor or study diseases, is referred to as *seroepidemiology* (Dorland, 2000).

The impact of seroepidemiological studies on designing new vaccination schemes and improving current policies has been demonstrated many times in the past. Using seroepidemiology, it is possible to estimate vaccine coverage when no other information is available, vaccine effectiveness (reduction in the incidence of a disease among individuals who have received the vaccine compared to the incidence in unvaccinated individuals), waning immunity over time or to identify susceptible cohorts (Giesecke J, 2002; Nelson KE & Masters Williams C, 2007). These estimates can be used to evaluate and compare

different vaccination strategies using mathematical modelling, in order to build more robust vaccination policies.

Examples of UK vaccine policies that have been influenced by serosurvey studies include the introduction of MMR in 1987, the measles and rubella (MR) “catch-up” campaign in 1994, the introduction of the second MMR vaccine dose in 1996 and the decisions not to implement universal immunisation programmes for HBV and VZV (Brisson M & Edmunds WJ, 2003; Farrington CP, 1990; Osborne K *et al.*, 2000).

The work carried out for this thesis is an extension of earlier work carried out as part of a large European seroepidemiological study. The aims and methods of that study are explained in brief below.

1.6 The European Sero-Epidemiology Network (ESEN) project

Many countries in Europe have serological surveillance schemes for evaluating their national vaccination programmes (Pinot de Moira A & Nardone A, 2005; Pebody RG *et al.*, 2007; Brisson M & Edmunds WJ, 2003; Meerhoff T *et al.*, 2004; Neal S & Efstratiou A, 2007). Coordination and harmonization of the national serological surveillance systems of vaccine preventable diseases in Europe could provide important information for tackling infectious diseases in the region, achieving the targets set by the WHO. This was the aim of the European Sero-Epidemiology Network (ESEN), a European project established in 1996 (Osborne K *et al.*, 1997).

Eight European countries (Denmark, Finland, France, Germany, Italy, Netherlands, Sweden and UK), as well as Australia, participated in the project and seroprevalence was estimated for the following infections: measles (de Melker H *et al.*, 2001), mumps (Nardone A *et al.*, 2003), rubella (Pebody RG *et al.*, 2000; Pebody RG *et al.*, 2001),

diphtheria (Edmunds WJ *et al.*, 2000;Edmunds WJ *et al.*, 2001) and pertussis (Pebody RG *et al.*, 2000). The project was co-coordinated by the Health Protection Agency (HPA), Centre for Infections (CfI) (previously known as Public Health Laboratory Service (PHLS)) based in Colindale, London.

1.7 Limitations of serosurveys

Two major limitations were identified and addressed before the beginning of the project. One obvious problem was each country's serosurvey representativeness. Collecting potentially thousands of specimens using random sampling had practical difficulties, especially for some of the smaller participant countries with insufficient resources and limited population size to sample from. Consequently, countries used different methods for collecting the specimens, mostly by utilising serum that was left over from specimens submitted to the laboratories for diagnostic purposes and others by carrying out population-based serosurveys (Osborne K *et al.*, 2000;Osborne K *et al.*, 1997).

A potentially more serious limitation was that direct comparisons of serological measurements between national serosurveys were not possible since each national laboratory tested their own country's samples using a variety of assays and techniques (Osborne K *et al.*, 2000). Different laboratory methods can introduce additional sources of variability, namely *between-laboratory* and *between-assay* variability and hence, non-comparable serological outcomes. The option of developing a set of guidelines for serological procedures (i.e. same assays and techniques) was not possible since many commercial assays were not available everywhere and laboratories preferred to use the techniques in which they had developed expertise.

1.8 The method of standardising serological results

In order to overcome the problem of between-laboratory and between-assay variability and achieve comparable outcomes, a method of standardising quantitative serological outcomes was developed (Osborne K *et al.*, 2000). For each infection, a reference laboratory tested a small panel of sera designed to cover negative, low and high positive results, and these were subsequently sent and tested by each country's national laboratory. The panel results for each country were regressed against the reference laboratory's results, and a standardisation equation was generated for each country that could transform national serosurvey results into common units. Once the serological results had been transformed into common units then a direct comparison between different countries' results was considered valid (Andrews N *et al.*, 2000; Giammanco A *et al.*, 2003; von Hunolstein C *et al.*, 2000).

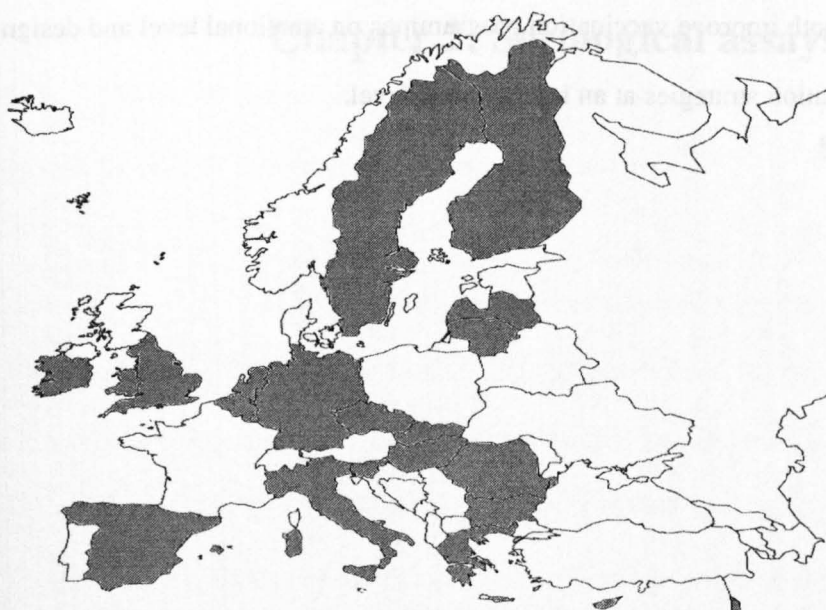
1.9 The European Sero-Epidemiology Network 2 (ESEN2) project

The European Sero-Epidemiology Network 2 (ESEN2) project, funded by the European Commission in 2001, was based on the original ESEN project. ESEN2 shared the same aim as the original project i.e. to standardise the serological surveillance of vaccine-preventable infections in order to improve vaccination policies across Europe.

An additional study objective of the ESEN2 project was the establishment of an active European network of experts in laboratory techniques and epidemiology for the surveillance of vaccine preventable infections. Moreover, the serological results, as mentioned earlier, apart from providing information with respect to the current risk of infection amongst the population could also be used for assessing the cost-effectiveness of vaccination programmes using mathematical modelling.

Apart from the infections included in the original ESEN project, the co-ordination of surveillance was extended to another three, namely VZV, HAV and HBV (eight infections in total). The number of participant countries was also increased to 22 (Figure 1.1) (including two countries outside the European region, Australia and Israel), in comparison to the 8 countries included in the original project (Nardone A & Miller E, 2004). Most countries did not test all the eight infections included in the ESEN2 project but only those that were considered to be of national public health interest.

Figure 1.1 A map of Europe highlighting the ESEN2 participant countries



1.10 Other international seroepidemiology studies

It is known that vaccination strategies can be more effective when coordinated under an orchestrated international effort. Examples of this are the immunisation campaigns against polio and smallpox that have succeeded in eliminating the former from the European region, and eradicating the latter worldwide (Nelson KE & Masters Williams C, 2007).

Unfortunately, there have been very few seroepidemiological projects on an international level to guide coordinating efforts with their findings. Some recent projects are an

examination of Parvovirus B19 infection in five European countries (Mossong J *et al.*, 2007), a prospective study of pregnant women of Human T-Lympotropic viruses between seven European countries (Taylor GP *et al.*, 2005) and an investigation of the seroepidemiology of herpes simplex virus type 1 and 2 in the population of eight European countries (Pebody RG *et al.*, 2007).

In conclusion, action at a national level may not be sufficient for the eradication or elimination of vaccine preventable diseases. Instead, a coordinated strategy at an international level is needed. Information gained from international seroepidemiological studies can be used to both improve vaccination programmes on a national level and design more efficient immunisation strategies at an international level.

Chapter 2: Serological assays, vaccines and seroprevalence

2.1 The immune system

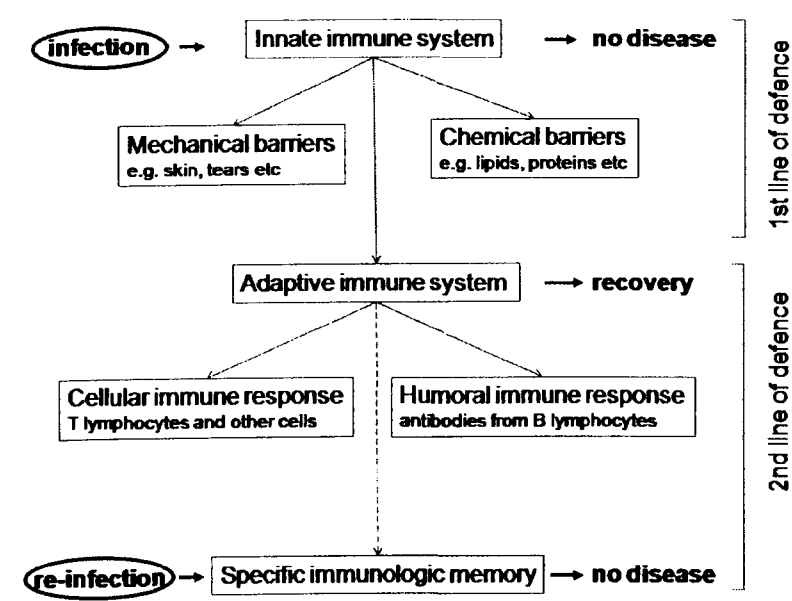
The body's *immune system* provides a defence against *pathogens* i.e. invasive organisms such as viruses, bacteria and fungi which may cause disease (Davies DH *et al.*, 1999;The Gale Group, 2008). The immune system can be divided into the *innate* or *non-specific immune system*, and the *adaptive* or *specific immune system*. The innate immune system is the “first line of defence” against infections and is designed to work within minutes after a pathogen attacks the body. It consists of *mechanical barriers* such as skin and mucous membranes, ciliated cells and mucus in the respiratory tract and the washing action of tears and urine. The innate immune system also consists of *chemical barriers*, such as lipids, lipoproteins and peptides in the skin, lysozyme in tears, several proteins in the oral cavity and the acidic pH in the stomach that are toxic to pathogens (Nelson KE & Masters Williams C, 2007).

The adaptive immune response - a more sophisticated immune response - is triggered once the innate immune system fails to eliminate the threat of infection. Unlike innate immunity which does not distinguish between pathogens, the adaptive immune system is designed to recognise and remove specific *antigens* (any (foreign) substance that generates anti-reaction) (Giesecke J, 2002). More specifically, receptors that are present on cells of the immune system can recognise small subregions of the antigen called *epitopes*. Different receptors in the adaptive immune system can recognise highly unique epitopes, and therefore, specific antigens. The cells of the adaptive immune system that are responsible for recognition of specific antigens are a type of leukocyte (Greek word for “white blood cells”), called lymphocytes, which are subdivided into T cells and B cells. When the receptors bind to the antigen, the B and T lymphocytes are activated, triggering the immune response (Nelson KE & Masters Williams C, 2007).

After the immune system has been activated and the antigen has been cleared, some B and T cells will become memory cells. These rapidly mobilise upon re-exposure with a previously encountered antigen (Nelson KE & Masters Williams C, 2007).

A simple description of the immune system response is shown in Figure 2.1. Initially, the pathogen encounters the innate immune system. This might prove sufficient to prevent infection. If the first line of defence fails to eliminate the infection, the adaptive immune system reacts with an antigen-specific response. In the case of re-infection, immunological memory to a specific antigen can be reactivated to produce a stronger and faster immune response (Mims C et al., 1998).

Figure 2.1 Innate and adaptive immune systems



Source: Based on a figure from “Medical Microbiology” (Mims C et al., 1998)

Adaptive immunity can be grouped into *cell-mediated* or *cellular immune response*, and *humoral immune response*. In the cellular immune response, T lymphocytes and other types of cells recognise and destroy antigens. The humoral immune response involves the

production of specific proteins from the cells of B lymphocytes called *antibodies* (or *immunoglobulins*) that bind to the antigen. This binding incapacitates the antigen and also stimulates removal of pathogens by macrophages and other cells (Giesecke J, 2002;Nelson KE & Masters Williams C, 2007).

There are five different varieties of antibodies (known as *isotypes*), each with different structures and roles: IgM, IgG, IgD, IgA, and IgE. IgM provides the majority of antigen-based immunity during the early stages of infection while IgG accounts for the majority of antibody-based immunity at a later stage of the infection (Nelson KE & Masters Williams C, 2007;Hardelid P, 2008).

2.2 Diagnostic tests

Biological assays (or more specifically, in this case, *serological assays*) are used to measure the existence, the amount and the type of antibodies in an individual's blood. However, before carrying out a serological test, the *serum* needs to be extracted from the blood i.e. the liquid that separates after the blood is allowed to completely clot (Saunders, 2007). The process of serum separation is achieved by leaving tubes of blood samples at room temperature for a while so they clot, and afterwards placing them in a *centrifuge machine*. A picture of samples placed in a centrifuge machine is given in Figure 2.2. After the samples have been centrifuged, the red blood cells, which are denser, settle at the bottom. The floating serum is then extracted into new tubes using a *Pasteur pipette* (Figure 2.3) (ProImmune Limited, 2009).

Figure 2.2 Test tubes in a centrifuge machine



Source: Wellcome images

Figure 2.3 Scientist uses a multi-channel Pasteur's pipette



Source: Wellcome images

Once the serum has been extracted, the immune status can be determined either by testing for a specific antigen or for antigen-specific antibodies. Moreover, assays can be designed

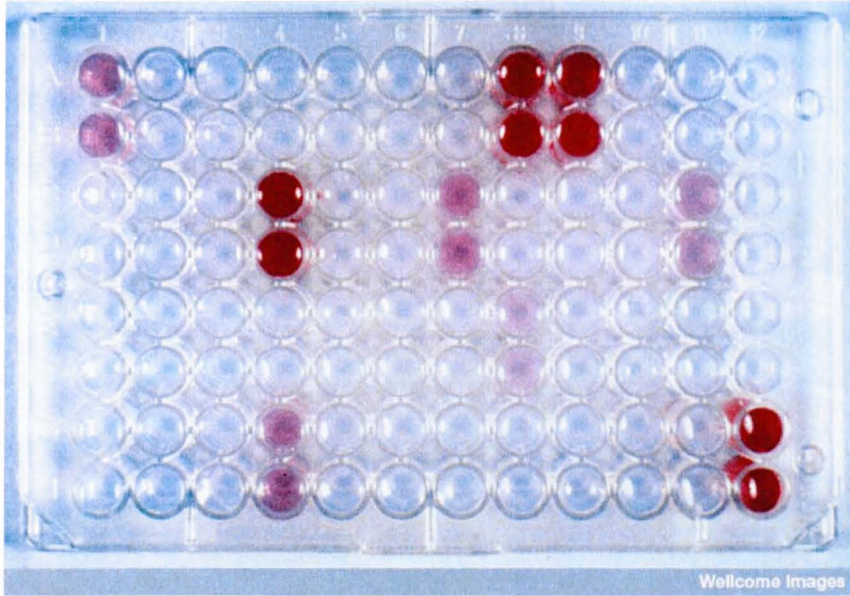
to test for a specific antibody isotype e.g. IgG or IgM (Nelson KE & Masters Williams C, 2007). Two commonly used assay methods are the enzyme-linked immunosorbent assay and the neutralisation test which are discussed briefly below.

(a) The enzyme-linked immunosorbent assay test

The *enzyme-linked immunosorbent assay (ELISA)* otherwise known as the *enzyme immunoassay (EIA)* is used to detect antibodies. Each well in a *microtitre* or *assay plate*, like the ones shown in Figure 2.4, is coated with a small amount of purified antigen. Once the serum has been extracted from the blood sample it is added to the well. If the serum contains any antibodies with specificity to the antigen in the well, they will bind to the antigen. The well is then washed and any material other than the bound antibodies is removed. The antibodies that are added next are linked to enzymes that can colour-flag the substance once a chromogenic enzyme substrate has been added. The resulting colour marks the quantity of antibodies present in the serum sample (Giesecke J, 2002).

Figure 2.4 shows an assay plate following an ELISA test. The coloured cells indicate a positive result (in this case chlamydial infection).

Figure 2.4 An ELISA plate with 96 wells



Source: Wellcome images

The difference in colour that indicates the concentration of antibody can be quantified using a *spectrophotometer* (Figure 2.5). The resulting measurement that is based on the the amount of light absorbed by the sample is called *absorbance* or *optical density (OD)* (Giesecke J, 2002).

Figure 2.5 Spectrophotometer for measuring the intensity of colour for an ELISA test



Source: Wellcome images

The OD measurements are transformed automatically to *International Units per millilitre (IU/ml)* by the assay reader using calibration equations. The quantity of antibodies is expressed by IU/ml. The OD transformation method and reasons why this is necessary are explained in detail in Chapter 9.

The method described above is a summary of a variety of techniques that use enzyme-linkage, and are all different variations of the ELISA assay method described above.

Antigens can be detected directly from the specimen, or once an organism is cultured outside the body i.e. *in vitro*. ELISA assays manufactured by a number of companies were used during the ESEN2 project.

(b) The neutralisation test

The *neutralisation test (NT)* is used for determining the presence or absence of pathogens. Serum samples are mixed with the virus and subsequently incubated in the presence of a susceptible cell type. If antigen-specific antibodies exist in the serum, they will bind to the

virus, thereby blocking viral infection in the susceptible cells. The level of infectivity is then measured (Nelson KE & Masters Williams C, 2007).

Since the NT is only able to detect the presence or absence of infection, quantification of such results is only possible with the use of *serial dilution*. The assay is initially conducted with the undiluted serum sample (dilution 1: 1). In the case of a “positive” result the sample is further diluted (i.e. dilution 1: 2). If the sample is still positive, further serial dilutions are carried out until a negative result is achieved. The highest dilution producing a positive reaction is called the *titre* (Giesecke J, 2002). For the ESEN2 project, the NT was the preferred assay by most laboratories to measure specific diphtheria toxin antibodies.

An example of the NT is shown in Table 2.1, where a sample was tested positive (dilution 1: 128), and was subsequently further diluted (dilution 1: 256). The titre is defined as 128 which is the inverse of the last dilution at which a positive result was observed.

Table 2.1 Example of serial dilution

Dilution	serum result
1: 8	+
1: 16	+
1: 32	+
1: 64	+
1: 128	+
1: 256	-

Although the titre takes the value of the highest positive dilution, in theory the sample may become negative at any time between the highest positive and the negative dilution. For the example given in Table 2.1, this would mean that the sample may have become negative at

any dilution level between the interval (128, 256] and therefore, the true titre lies in between these two values. Although in practice this is seldom a problem, an example in which an adjustment was made to specify the true titre is presented in Chapter 5.

Serial dilutions can quantify binary assay results resulting in a “semi-quantitative” format. An example of a frequency distribution is given in Table 2.2 where the Luxembourg national laboratory tested 146 sera against diphtheria antigen using the NT assay. Note that the measurements shown are not from a random population sample but were selected in a way to include measurements ranging from negative to high positive.

Table 2.2 Distribution of 146 sera from Luxembourg tested against diphtheria using the NT assay

Titres	frequency
8	43
16	15
32	8
64	18
128	21
256	19
512	11
1024	8
2048	1
4096	1
8192	1

Due to the exponential nature of serological data, it is common practice to transform such results using logarithms. A logarithmic transformation of base 10 was applied to all serological data for the ESEN2 project, and this was also the practice for the analysis presented in this thesis.

2.3 Vaccines

To be effective, a vaccine needs to produce considerable humoral and cell-mediated immune responses and provide protection for several years; if not for a lifetime.

Essentially there are two main types of vaccines: (a) Live organisms that have been artificially weakened (*attenuated*) and (b) *inactive* organisms that are either whole or fractioned.

(a) Live attenuated vaccines

Attenuated bacteria and viruses are infective but non-pathogenic, and are produced by culturing pathogenic strains in vitro or in unnatural hosts, aiming to discover strains with restricted virulence growth in humans. Live attenuated vaccines provide strong immunity that tends to persist for a long time after vaccination, due to immunological memory, and have the advantage of producing both humoral and cell-mediated-immunity. A disadvantage of such vaccines is the likelihood of adverse effects, since they are poorly tolerated by immunocompromised individuals. In addition, live vaccines must be carefully stored and refrigerated to maintain their activity. Examples of live attenuated vaccines are measles, mumps and rubella (Nelson KE & Masters Williams C, 2007; Davies DH *et al.*, 1999).

(b) Inactivated vaccines

Inactivated vaccines use viruses and bacteria that have been killed by heat or chemicals. Otherwise inactivated vaccines may use purified components of the organisms (Nelson KE & Masters Williams C, 2007). The advantage of these vaccines is that they don't need to be refrigerated. The disadvantage is that they do not produce an antibody response as strong or as long lasting as live vaccines. As a result, several booster doses may be needed

to reinforce the antibody response (Davies DH *et al.*, 1999). An example of an inactivated vaccine is the hepatitis A vaccine (Nelson KE & Masters Williams C, 2007).

Apart from the traditional vaccine types described above, over the last two decades there has been much interest in a new type of vaccine, the *recombinant vaccine*. This involves the genetic manipulation of organisms to create either active or inactive vaccines (Nelson KE & Masters Williams C, 2007).

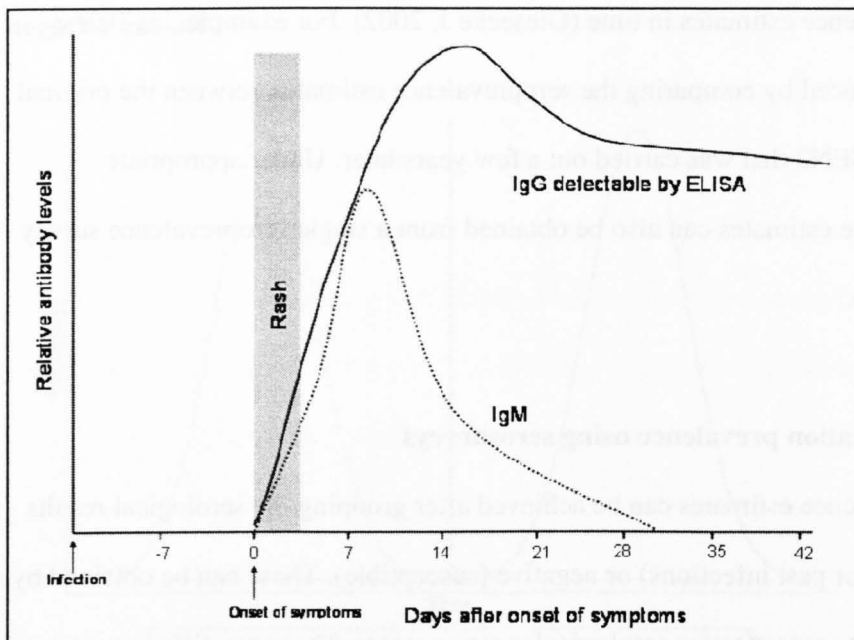
2.4 Prevalence and incidence definitions

As defined in the previous Chapter, prevalence is the number (or proportion) of current cases of a disease in a given population at a specific time (Mosby, 2009). Seroprevalence, on the other hand, is the number (or more usually, the proportion) of individuals with a serological marker of infection at a specific time. Thus, seroprevalence is defined like prevalence, but its interpretation in terms of disease or infection can be quite different, depending on the marker. The serological markers considered in this thesis are quantitative measurements that have been dichotomised as positive or negative.

As mentioned earlier, IgM antibody type appears in the serum for a limited period of time following the infection, and therefore, its presence indicates recent infection. IgG antibody is used as evidence of recent or past infection since it sometime peaks later but remains in the serum for a long period of time (Giesecke J, 2002).

Figure 2.6 shows the rubella IgG and IgM levels after the onset of symptoms. Both IgG and IgM levels rise about the same time but IgM levels rapidly decrease to their previous levels after approximately a month. Therefore, a positive IgM provides evidence of recent (or current) infection, whereas a positive IgG can either mean current or past infection.

Figure 2.6 Relative rubella IgG and IgM levels after onset of rubella symptoms



Source: PhD thesis titled “The seroepidemiology of rubella in pregnant women in North Thames” (Hardelid P, 2008)

For the antigens included in the ESEN2 project the serosurveys were tested against IgG specific antibodies. An exception was HBV which is associated with a number of different antigens (Mims C *et al.*, 1998). Samples were tested for antibodies against Hepatitis B core antigen (anti-HBc), Hepatitis B surface antigen (anti-HBs) as well as Hepatitis B surface antigen (HBsAg). Interpretation of these assays is more complex and is beyond the scope of this project.

As mentioned earlier, for many infections, naturally-induced high IgG antibody levels remain in the blood for many years after the infection, waning little with time, and for many pathogens, natural infection can be assumed to provide lifelong immunity (Vyse AJ *et al.*, 2006). Therefore, given this assumption is valid, testing for IgG antibodies results in seroprevalences that cover both recent and past infections.

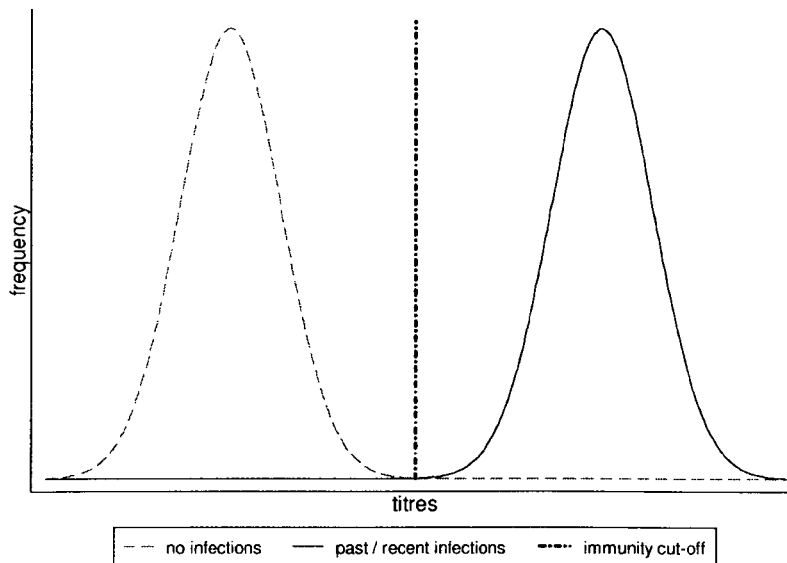
Incidence and prevalence (and seroprevalence) are related. Incidence can be estimated by comparing seroprevalence estimates in time (Giesecke J, 2002). For example, incidence estimates can be produced by comparing the seroprevalence estimates between the original ESEN project, and ESEN2 that was carried out a few years later. Under appropriate assumptions, incidence estimates can also be obtained from a single seroprevalence survey (Farrington CP, 1990).

2.5 Estimating population prevalence using serosurveys

Population seroprevalence estimates can be achieved after grouping the serological results into positive (current or past infections) or negative (susceptible). These can be obtained by applying a cut-off to the quantitative serological measurements. There are different methods for estimating such cut-offs, and well-conducted seroepidemiological studies are expected to include details of how they were calculated (Giesecke J, 2002).

The example in Figure 2.7 shows the distribution of antibody titres as separate components, one corresponding to current / past infections, and another to non-infected individuals. These distributions are only hypothetical since the true status of the individuals is unknown. Serological measurements higher than the cut-off value were classified as positive whereas those lower as negative.

Figure 2.7 Titre distributions for infected and non-infected individuals and positive / negative cut-off



Mixture modelling is a method that has been used in the past to estimate population prevalence from serological data (Gay NJ, 1996). The principle is based on a mixture of component distributions for distinct underlying population groups. Using this method it is possible to distinguish between high IgG levels attributed to current or past infections, and low antibody levels attributed to non-infected and therefore, susceptible individuals. The proportion of individuals falling in each component can then be estimated. Using the same example as above (Figure 2.7), the proportion of past / recent infections will be equivalent to seroprevalence. Mixture models will be discussed in detail in Chapter 7.

2.6 Estimating seroprevalence for a population under an existing vaccination programme

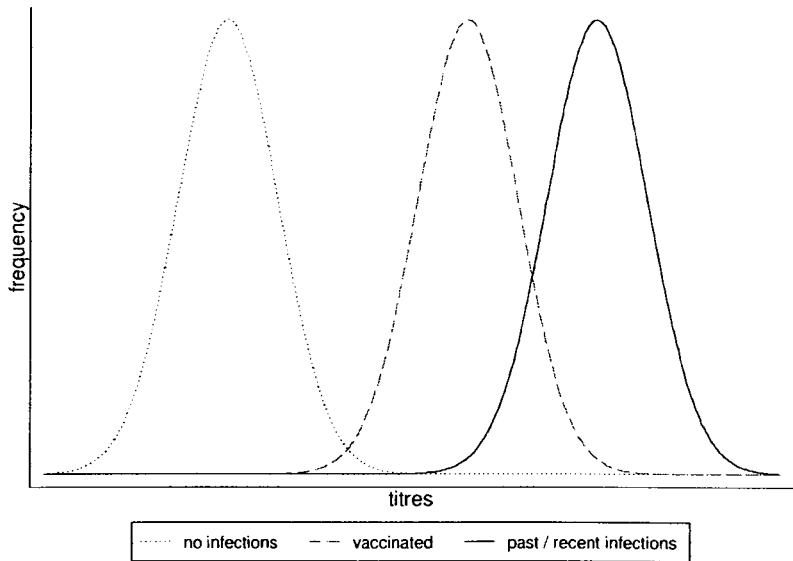
Estimation of seroprevalence becomes more complex when part of the population has been vaccinated. This is subject to several factors such as the vaccine type (e.g. live or inactivated), the vaccine dose, the individuals' characteristics (age, socioeconomic status, state of the immune system, genetic factors etc) and the vaccine schedule (timing and

number of doses). Antibody responses to vaccination are often lower than those resulting from natural infection. Moreover, vaccine-induced IgG levels may wane with time, with the degree of decrease varying greatly between different vaccine types and antigens (Giesecke J, 2002; Vyse AJ *et al.*, 2006). Potentially, this means that part of the vaccinated population may become susceptible to an infectious disease after a period of time. Outbreaks among populations with high vaccine coverage that were previously considered to be protected confirm this. Such examples include the recent mumps outbreak among the Jewish community of New York, the pertussis outbreak in the Netherlands and the diphtheria outbreak in the ex-Soviet Union states during the 1990s (de Melker H *et al.*, 1997; Hardy IR *et al.*, 1996; Centers for Disease Control and Prevention, 2010).

Figure 2.8 depicts three hypothetical titre component distributions for a population with an existing vaccination programme. Between the components of past / current infections and non-infected individuals, a third component reflects the group of vaccinated individuals.

As before, mixture modelling allows for an estimation of the proportion of individuals falling into each component. In this case a seroprevalence definition would include both groups of current / past infections and vaccinated individuals.

Figure 2.8 Titre distributions for infected, vaccinated and non-infected / non-vaccinated individuals



2.7 Correlates of protection

The measurable quantities in the immune system that correlate with protection against an infection or disease are called *Correlates of Protection (CoP)*. The identification of such proxy measures is significant for vaccine development, since they are employed to estimate the proportion of *seroconverters* as a result of vaccination (i.e. the proportion of a vaccinated group that acquires protection after vaccination), and hence, the reduction in the disease incidence following vaccination, called *vaccine efficacy (VE)* (Note the difference between the terms VE and vaccine effectiveness mentioned earlier; the former is estimated from clinical trials and therefore does not include the vaccine's indirect effect i.e. its ability to reduce the spread of infection in the population). Often vaccines have excellent efficacy records based on empirical observations, without a clear understanding of the immune system mechanisms of protection (Lambert PH *et al.*, 2005; Plikaytis BD & Carlone GM, 2005; Qin L *et al.*, 2007).

For the diagnostic assays included in the ESEN2 project (mainly ELISA and NT), antibodies were used as CoP. Research in the past has shown that for most (but not all) vaccines, prevention of infection correlates well with specific antibody response. An example of such a study was published in 1990, including schoolchildren tested for measles prior to an outbreak. It was shown that in general, antibody titres above a specific value correspond to protection, although occasionally cases occur in individuals who have previously been infected (Chen RT *et al.*, 1990). For some infections there are no known serological correlates of protection. This is arguably the case, for example, for pertussis infection.

The relationship between protection and antibody production is often more complex than simply the level of serum antibodies. For example, it may be important for antibodies to be present at the site of virus replication, or have the ability to affect multiple strains of a virus or viruses that may potentially mutate (Plotkin SA, 2010). Moreover, immunological memory should be generated in such a way that in the case of infection recurrence a sufficient number of antibodies will concentrate rapidly around the site of the pathogen (Lambert PH *et al.*, 2005).

In conclusion, one of the assumptions used in the ESEN2 project was that of a high correlation between antibody response and protection. This assumption will be assumed to hold for this thesis hereafter. Although this assumption is believed to generally hold for the antigens included in the project (with the exception of pertussis), the relationship between antibody response and protection may not be so strong for some antigens such as mumps where outbreaks have been recorded for highly immunised populations (Centers for Disease Control and Prevention, 2010).

2.8 Seroprevalence is not always equivalent to proportion protected

As mentioned in the previous Chapter, the main aim of the ESEN2 project was to identify susceptible cohorts in the population, or inversely, to recognise the proportion of population “protected” against a particular disease. For an unvaccinated population the proportion of past / recent infections should equal seroprevalence, assuming life-long immunity following infection. However, there are a number of different factors that can affect the performance of diagnostic assays, which means that seropositivity and protection may not always be equivalent.

The performance of an assay test can be assessed by calculating *sensitivity* and *specificity*. Sensitivity is defined as the proportion of individuals that have high antibodies to be recognised as such by the test, whereas specificity is the proportion of individuals that have low antibodies to be recognised as such by the test (Ades AE, 1990). By re-estimating the assay cut-off, either of these quantities can be increased at the expense of the other i.e. it is not possible to improve both sensitivity and specificity simultaneously. A key issue on defining an assay cut-off that achieves an optimal balance between sensitivity and specificity depends on whether the assay is designed to be used for individual diagnosis or for epidemiological purposes. This will be discussed in detail in Chapter 8.

Apart from the cut-off definition there are other factors that determine immunity which the marker may not capture. Antibody *cross-reactivity* is defined as the ability of antibodies to react with more than one antigen (Male D *et al.*, 2006). Cross-reactivity can become a problem with respect to diagnostic assays, since it can produce evidence of antibody reaction, even when little or no antibodies exist for a specific pathogen. Although cross-reactivity can be evaluated for assays such as ELISA, it may be possible that it is an issue for some diagnostic tests and antigens included in the ESEN2 project (Ndumbe PM & Lvenisky RJ, 1985).

Finally, previous studies have shown that antibody levels wane with time, depending on the antigen and whether the response is induced by natural infection or vaccination. As mentioned earlier for natural infections, antibody levels stay generally stable with time, however vaccine-induced immunity often wanes rapidly (Vyse AJ *et al.*, 2006). This can produce a number of measurements that are borderline between negative and positive, without it being possible to distinguish whether these individuals are truly protected.

Chapter 3: Aims and objectives

3.1 PhD project rationale

This thesis is largely based on the ESEN2 project, from where it drew its general aim and material. However, it needs to be viewed as an extension of the ESEN2 project, since - with the exception of the formalisation of the standardisation methodology shown in Chapter 4 that was used during the project - the work carried out for this thesis was beyond the scope of the project.

As was also mentioned in Chapter 1, it is possible to identify susceptible groups within the population by using serological data. In the short term, these groups can be specifically targeted by booster campaigns so as to prevent future outbreaks and limit the spread of the infection. In the long term, serological information can be used to evaluate and improve vaccination programmes. International studies based on population serological data allow for comparisons of the levels of immunity in different populations to be made along with the assessment of different vaccination strategies.

As mentioned in Chapter 1, the ESEN2 project aimed to achieve comparable seroprevalence estimates across Europe. Validating the methods used during the ESEN2 project for harmonising serological outcomes is a crucial aspect of this overarching aim. Furthermore, it is important to compare the method employed for ESEN2 with alternative seroprevalence estimation techniques aiming to achieve improved methods that could potentially be used in future seroepidemiological studies.

3.2 Overall aim

The overall aim of this thesis was to formalise, validate and further develop the methodology used for the ESEN2 project for harmonising serological results, and to propose alternative methods for achieving comparable population seroprevalence estimates across Europe.

3.3 Specific objectives

Specific objectives of this work are the following:

1. To develop a statistical-based algorithm for implementing the standardisation methodology.
2. To further develop the standardisation algorithm by taking into account issues such as censored data.
3. To assess the impact of the standardisation method on seroprevalence.
4. To examine and quantify the sources of variability arising from laboratory testing such as the between-laboratory and between-assay variability.
5. To compare the standardisation methodology with an alternative method of seroprevalence estimation (mixture modelling).
6. To assess the effect of the assay cut-offs on seroprevalence estimation and to suggest methods for re-estimating cut-offs suitable for seroepidemiology.
7. To investigate the impact of plate-to-plate variability on seroprevalence estimation.

3.4 Statistical methods

A number of statistical techniques have been used throughout the thesis. However, their implementation will not be presented in a separate “methods” Chapter. Instead each statistical tool will be introduced when appropriate.

A list of the statistical tools introduced in the next Chapters is given below:

Censored regression

Different types of censored regression models were used to obtain standardisation equations as an alternative to the regression models used for the ESEN2 project (Chapter 5).

Multiple imputations

Standardisation equations were obtained from models using a method of multiple imputations to account for censored data (Chapter 5).

Non-parametric test for censored data

A non-parametric test was used to indicate whether standardisation was valid to use in certain cases of censored serological data (Chapter 5).

Mixture models for seroprevalence estimation

Different types of mixture models were used, varying the distribution of the underlying components, to estimate seroprevalence from the non-standardised serosurvey results (Chapter 7).

Mixture models for cut-off estimation

A method was developed based on mixture models that included cut-off re-estimation for obtaining seroprevalences (Chapter 8).

Mixture models for investigating plate-to-plate variability

Different methods based on mixture modelling were used to investigate and adjust the effect of plate-to-plate variability on population seroprevalence (Chapter 9).

A summary of the Chapters to follow, along with a description of how the statistical methods described above were used, is given in the following Sections.

3.5 Developing further the standardisation methodology

As explained in Chapter 1, the method of standardisation was used in the original ESEN project to transform each country's quantitative serological results into a comparable outcome, based on standardisation equations derived from testing a panel of sera by each national laboratory (Andrews N *et al.*, 2000; Nardone A & Miller E, 2004).

As part of this thesis, the standardisation methodology used for the original ESEN project to account for issues such as *outliers*, *censored data* and selecting between different types of *standardisation equations* was formalised, and a statistical-based algorithm was established (Chapter 4). This algorithm was subsequently used throughout the ESEN2 project for standardising the serological results into common units.

One aspect of the methodology that was further investigated was the effect of censored observations on the standardisation equation. For many serological assays no quantitative results were reported if they were outside the assay *detection limits*. Such observations were treated as censored and were taken into account during the ESEN2 project analysis using a method of *simple substitution*. The validity of the simple substitution method, as well as other methods such as *censored regression* and *multiple imputation*, were investigated using simulations. A non-parametric test was proposed for censored data on the *x*-axis (Chapter 5).

3.6 Impact of standardisation on seroprevalence

Once a standardisation equation had been selected, it was used to transform the serosurvey data tested within each national laboratory into common units as defined by the reference laboratory assay. The transformation of the serological results was a way to correct for the between-laboratory and between-assay variability. The impact and scale of the different sources of variability on national seroprevalence were assessed by comparing the standardised with the non-standardised estimates (Chapter 6).

3.7 Mixture models

An alternative method for estimating population seroprevalence is the application of *mixture models*. Using this method, the standardisation of serological results can be bypassed altogether since mixture models can be applied directly on the non-standardised results. Several types of mixture models were fitted, and a comparison between standardisation and mixture model seroprevalence estimate results was carried out. The advantages and disadvantages of each method are discussed in Chapter 7.

3.8 Assay cut-offs

For the ESEN2 project, once the serosurvey results had been standardised, they were classified into positive or negative according to a cut-off point. As a consequence of this, the definition of the assay cut-off was very important, since it influences the classification of samples, and therefore, the estimated national seroprevalence. For the purposes of the ESEN2 project, the assay cut-off as defined by the assay manufacturer of the reference centre was used. In Chapter 8, the validity and impact of these cut-offs on the national seroprevalence are examined. Alternative methods for cut-off estimation are also examined based on ROC curve analysis and mixture modelling.

3.9 Plate-to-plate variability

Using standardisation it is possible to adjust for between-laboratory and between-assay variability. However, a certain amount of variability would still be present, even if samples were tested using a unique assay within the same laboratory. Some of the unexplained variability could be attributed to the fact that not all the samples were tested simultaneously but batches of them were placed and tested on different plates. In Chapter 9, the methods currently used for adjusting for *plate-to-plate* variability are validated. Also their impact on estimating population seroprevalence is examined and a method to adjust for this variability using mixture modelling is shown.

3.10 Statistical software

The main statistical software that was used for the data manipulation and analysis was Stata (StataCorp. Stata statistical software: releases 10.1, 11.0 and 11.1. College Station, Tex.: Stata Corporation, 2001). The censored regression models, the non-linear regression equations and the mixture models were all fitted using the maximum likelihood estimation command `ml` in Stata. Some general types of censored regression models programmed in Stata are given in Appendix I(A). An example of a programme for multiple imputations is shown in Appendix I(B), whereas a mixture model example is given in Appendix I(C). Note that Stata uses the Newton-Raphson algorithm for optimisation.

The `optimize` command in R statistical software was used for cut-off estimation in Chapter 8 (The R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing. Version: 2.10.1) (Brent R, 1973). An example of a cut-off estimation programme in R is given in Appendix I(D).

For all the models fitted and presented in this thesis a number of different starting values were tried, in order to check the validity of the convergence. Despite changing the starting values, the models fitted generally converged to the same point which supports the robustness of the maximization procedure. Prior to fitting complex models such as the mixture model shown in Chapters 7, 8 and 9, simpler, more robust models were fitted and their estimates used as starting values for the more complicated models.

In addition, the same models were fitted using alternative commands or software to check the model fit. For example, linear censored regression models were fitted using both the `ml` command and the built-in `cnreg` command in Stata (part of the official software up to version 11.1).

All figures included in the thesis were constructed using Stata, whereas Microsoft Excel was used for the tables.

3.11 Data source

The data that are presented throughout this thesis were collected as part of the ESEN2 project. Permission has been granted to the author by the members of the ESEN2 group to use this data.

3.12 Publications arising from the thesis

The author started working on the ESEN2 project in January 2002 (the project officially started in 2001), whereas registration for the PhD thesis commenced two years later, in January 2004. The work carried out for the purposes of this thesis is closely related to that of ESEN2. Although the methodology for the ESEN2 project was mainly defined at the beginning of the project, subsequent work was influenced in some degree by the findings of this PhD thesis.

A peer-reviewed paper was published based on material presented in Chapter 2. Two further papers were published as part of the ESEN2 project, but were influenced by work carried out for this thesis (Appendix II). Parts of the thesis were also presented in a number of conferences and seminars.

Chapter 4: Standardisation of serological results

4.1 Comparability of national seroprevalence estimates

The ESEN project aimed to compare and assess different vaccination strategies across Europe by identifying susceptible cohorts. In order to achieve this, age-specific seroprevalence needed to be estimated for each country which meant that comparable serological results were essential.

Although there was an effort to use the same assays at different national laboratories, this was not possible in practice because commercial assays were not available everywhere, and laboratories preferred to use assays they were most familiar with. Given the variety of assays used in the end, the resulting serological outcomes were not comparable between national laboratories. Furthermore, even in cases where the same assays were used the results were often still non-comparable due to differences in laboratory techniques and operating procedures.

One solution that would have enabled direct comparisons among serological results from the different national serosurveys, would be to test all samples in the same laboratory. Unfortunately, this was not possible during the ESEN projects given that large numbers of samples that needed to be tested. Moreover, one of the objectives of the projects was to develop a laboratory and epidemiology network of experts. As a result, it was necessary to develop a method for standardising serological results.

The algorithm described in this Chapter was used to standardise the ESEN2 serological results and is based on the methods initially used for the original ESEN project (Andrews *N et al.*, 2000). The algorithm takes into account issues such as the type of equation to be selected for standardisation, extreme observations that need to be investigated and results outside the quantitative range of the assay (Kafatos G *et al.*, 2005).

4.2 An example comparing seroprevalence

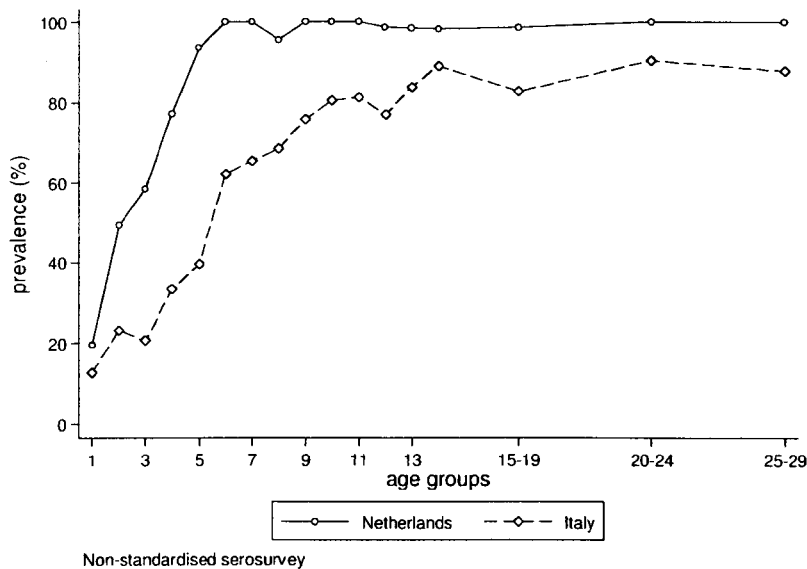
An example is given to demonstrate the incompatibility of the serological results (and hence, of the seroprevalence estimates) arising from the differing laboratory methods. Age-specific seroprevalence for VZV was estimated for Italy and the Netherlands after their laboratories used assay methods of their choice to test the national serosurveys.

In the Netherlands, the serosurvey conducted in 1998 included the collection of 1176 age-stratified samples. The assay method used was Human (Instruchemie bv). The resulting quantitative measurements were classified into positive or negative according to a cut-off of 0.3 IU/ml as set up by the assay manufacturer. The Italian national serosurvey consisted of 2446 age-stratified samples collected in 1996. An Enzygnost assay was used with assay cut-off 0.05 IU/ml.

After testing the samples and grouping the resulting quantitative measurements into positives and negatives, the seroprevalence of VZV was estimated for each age group.

Figure 4.1 shows the age-specific seroprevalence for Italy and the Netherlands. The seroprevalence for Italy was estimated to be much lower than the Netherlands for all age groups. However, it is unclear whether the difference observed represents a true seroprevalence difference resulting from genetic or environmental factors or whether it is a result of the different assay methods used.

Figure 4.1 Age-specific seroprevalence estimates of VZV antibody for the Netherlands and Italy (non-standardised)



A way to overcome the issue of non-comparable outcomes is to standardise quantitative serological results into common units. The method consists of constructing a small panel of sera by a reference country that can then be sent and tested by each national laboratory. Based on the panel results, regression equations can be selected to standardise each national serosurvey's results into common units.

4.3 Panel construction

At the start of the ESEN2 project, for each antigen, a reference laboratory prepared a standardisation panel consisting of approximately 150 samples. There were no formal sample size calculations for this, but it was chosen in a way to ensure that there were enough samples throughout the quantitative range of the assay from negative and equivocal (or low positive), to positive samples. Although a smaller sample size would probably be sufficient for most assays, a larger sample size was used to ensure there were enough

samples even for poorer assays with high variability. A list of the reference laboratories for each antigen is given in Table 4.1.

Table 4.1 Reference centres and assay methods used for testing the standardisation panels for each antigen

Antigen	Reference laboratory	Country	Assay method
Measles	Robert Koch Institute (RKI)	Germany	Enzygnost ELISA
Mumps	Robert Koch Institute (RKI)	Germany	Enzygnost ELISA
Rubella	Robert Koch Institute (RKI)	Germany	Enzygnost ELISA
Pertussis	University of Palermo (UoP)	Italy	In-House, ELISA
Diphtheria	Istituto Superiore di Sanità (ISS)	Italy	DA-DELPIA
Varicella Zoster Virus	Instituto de Salud Carlos III (ISCIII)	Spain	Enzygnost ELISA
Hepatitis A virus	National Retrovirus Reference Centre	Greece	Abbott (AxSYM)
Hepatitis B virus	National Retrovirus Reference Centre	Greece	AxSYM AUSAB

The reference laboratory tested the panel at least twice, and the geometric mean of these test results was taken to reduce the test-to-test variability for the reference centre. The panel was subsequently sent to each participant laboratory where it was tested by the assay method of their choice (Andrews N *et al.*, 2000).

4.4 Comparison of first and second panel test results

The standardisation panels were also tested twice by the participant laboratories. The aim of the first test that was carried out at the beginning of the project was to ensure that all assays performed adequately. Any major problems with the assay methods used were detected and amended at this stage of the project.

The panel was tested for a second time at a later stage, halfway through testing the serosurvey, and these second panel results were used for the selection of the standardisation equation.

The two test results allowed potential changes of the serological measurements over time to be investigated. This was examined using the following three methods:

(a) A scatter-plot was produced to investigate the relationship between the two sets of results.

(b) A paired t -test was used to test for the presence of a systematic bias.

(c) A Bland-Altman plot was constructed to show potential systematic bias and identify possible outliers. The differences between the two test results were plotted against their means following logarithmic transformation. Assuming the mean difference between the test results is estimated as $\hat{\mu}_d$ and the standard deviation of the differences as $\hat{\sigma}_d$, we would expect most differences to lie between the interval $\hat{\mu}_d - 1.96\hat{\sigma}_d$ and $\hat{\mu}_d + 1.96\hat{\sigma}_d$ (Bland JM & Altman DG, 1986).

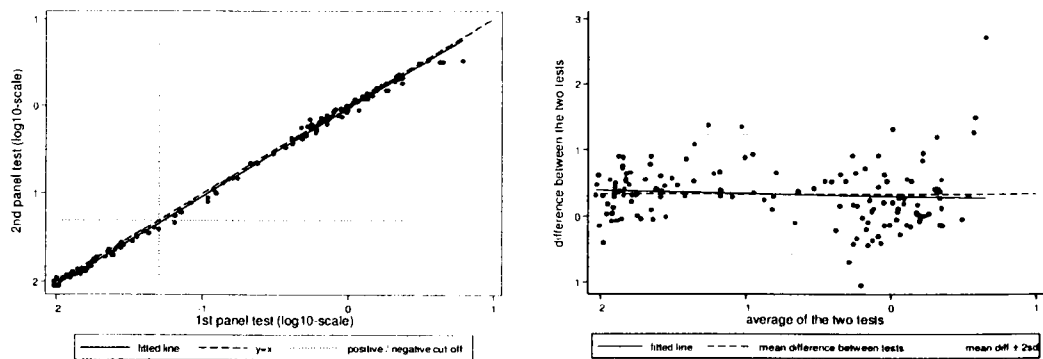
In an example comparing the first and the second VZV panel tests for Finland, the slope estimate of the regression line was 0.99 with a 95% *confidence interval (CI)* of [0.96, 1.02] which includes the line of equivalence. The Bland-Altman plot showed that the mean difference was close to zero, so there was little evidence of overall bias (figure 4.2a).

Figure 4.2b shows an example of HAV panel test results for Belgium, where the difference between the two tests was found to be significant. The regression slope did not include the line of equivalence (estimate = 1.03; 95% CI: [1.01, 1.04]). The Bland-Altman test showed that the second test gave consistently higher measurements than the first, and also that the

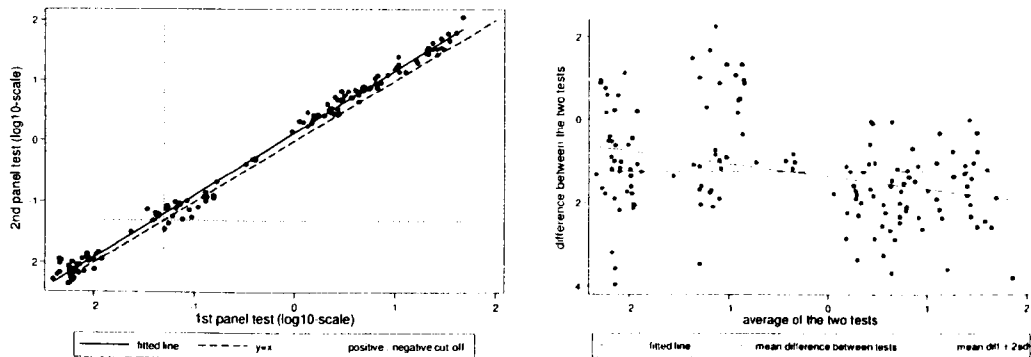
differences seem to increase for high values. The regression slope for the Bland-Altman test was estimated as -0.03 (95% CI: $[-0.04, -0.02]$).

Figure 4.2 Comparison between first and second panel tests

(a) Regression and Bland-Altman plots for VZV panel test results for Finland



(b) Regression and Bland-Altman plots for HAV panel test results for Belgium



The analysis comparing first and second tests of the panel often showed statistically significant differences as it was shown in the HAV example above. However, unless there had been a change in assay methodology, or a problem with the first test of the panel, these differences were usually small and reflected normal run-to-run assay variability. Such differences would not greatly affect the standardised results. In case of the differences being large and there being no clear problems with the first test, a third testing of the panel was undertaken.

4.5 Selecting standardisation equations

Once the second panel was tested by both the reference centre and the testing laboratory, their \log_{10} -transformed results were regressed (testing laboratory results were regressed against the reference laboratory results), and an equation was produced to be used for standardising the serosurvey results.

A polynomial regression equation of the linear form $g(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k$ can be defined as:

$$y_i = g(x_i) + \varepsilon_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k + \varepsilon_i ,$$

Equation 4.1

where $i = 1, \dots, n$ index the sampled observations, y_i is the response variable, x_i the explanatory variable and ε_i the error term for $k + 1$ parameters. The error is assumed to be identically and independently normally distributed with zero mean and constant variance σ^2 i.e. $\varepsilon_i \sim N(0, \sigma^2)$.

The fitted values are given by:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 + \dots + \hat{\beta}_k x_i^k ,$$

Equation 4.2

where $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ are the parameter estimates estimated by maximum likelihood.

The *residuals* $\hat{\varepsilon}_i$ are defined as the deviations from the fitted values i.e.

$$\hat{\epsilon}_i = y_i - \hat{y}_i$$

Equation 4.3

(Fox J, 1991).

As an example, the VZV panel was tested by both the Belgium and the Spanish reference laboratory using Enzygnost ELISA assays. The Belgian results were regressed against the Spanish data using a quadratic model determined by the prediction equation of the form:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 x_i^2 ,$$

Equation 4.4

which in this example becomes:

$$\hat{y}_i = -0.3 + x_i + 0.1x_i^2 ,$$

where $i = 1, \dots, 119$, x_i are the \log_{10} -transformed measurements given by the reference laboratory and \hat{y}_i are the predicted \log_{10} -transformed Belgian results.

A number of issues needed to be considered for estimating standardisation equations.

These are the following:

(a) The type of equation selected. In the example above a quadratic equation was used but other types may be considered, which means that a robust selection method is necessary.

(b) Extreme observations need to be flagged as *outliers* and corrected if wrong (e.g. due to errors occurring during data entry or laboratory procedures) or otherwise excluded from further analysis.

(c) An issue common with serological assays is the existence of a detection range within which observations are given in a quantitative form. Any measurements outside this range, say (D_L, D_U) , are depicted as “ $< D_L$ ” or “ $> D_U$ ” if they are below or above, respectively.

For example, the Spanish laboratory tested their panel for VZV using an assay with detection range between 0.002 and 6.62 IU/ml. This meant that quantitative results were returned for 124 samples (85%), 19 samples (13%) were recorded as “ < 0.002 ” and 5 (3%) as “ > 6.62 IU/ml”.

The following three Sections explain in detail how these issues were taken into account in the analysis that was carried out as part of the ESEN2 project.

4.6 Outlier investigation

After plotting each country’s outcomes against the reference laboratory’s results, one or more observations were sometimes considered extreme compared to the rest of the data. These potential outliers were investigated, and any errors that occurred during the panel testing, data entry or data manipulation were corrected. If any extreme measurements could not be attributed to any obvious error, it was assumed that an error had occurred during the laboratory testing, and they were omitted from further analysis. A statistical procedure was defined to decide which results had to be excluded.

Once a linear regression equation of the type shown in Equation 4.1 was fitted, the *standardised residuals* z_i were calculated as shown in Equation 4.5.

$$z_i = \frac{\hat{\epsilon}_i}{\hat{\sigma}},$$

Equation 4.5

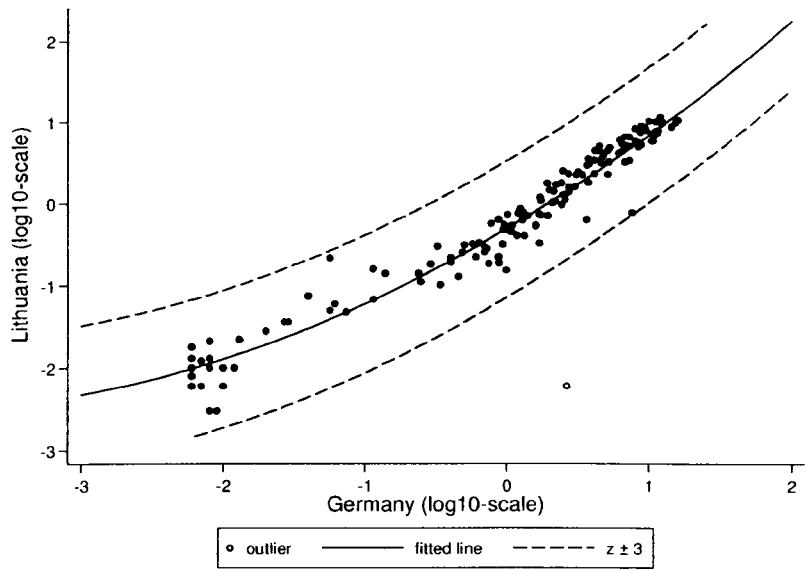
where $\hat{\sigma}$ is the maximum likelihood estimate of the standard deviation σ (Armitage P & Berry G, 1994).

Any observations with z_i outside the range of ± 3 (approximately 1% of the data) were considered outliers and were excluded from further analysis. This step was repeated with a more relaxed exclusion criterion of ± 4 . A two-step analysis was chosen to avoid a high number of iterations. A wider range was used on the second iteration to minimise the risk of too many samples being dropped as outliers. The maximum number of points dropped as outliers was set as 5% of the total samples. In cases where more points were flagged as outliers, the more extreme 5% of the data were dropped. However, such a large number of outliers did not occur in any of the panel comparisons on the ESEN2 project.

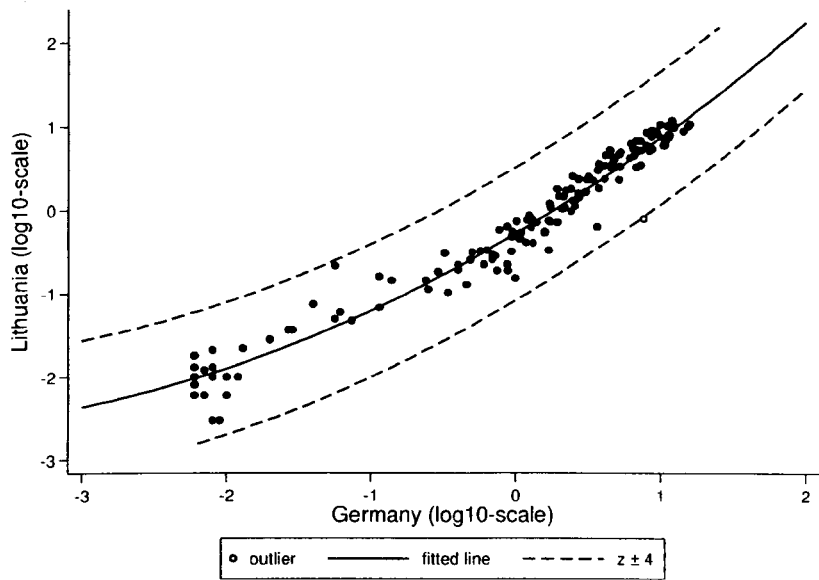
An example using the measles sample results of Lithuania regressed against the German reference centre's results is given (both laboratories used Enzygnost ELISA assays). A quadratic equation was initially fitted and the standardised residuals were calculated. The standardised residual was outside the ± 3 range for one observation and is highlighted in Figure 4.3(a). This observation was omitted and the model was re-fitted. Using a wider criterion for the standardised residuals of ± 4 , there was one additional observation that was borderline defined as outlier and was also omitted (highlighted in Figure 4.3(b)).

Figure 4.3 Outlier investigation for Lithuanian measles results regressed against the Germany reference centre

(a) Standardised residuals outside the range ± 3



(b) Standardised residuals outside the range ± 4 (after the model was re-fitted)



4.7 Selection of regression model

When selecting a regression model to be used for standardisation, the main aim was to choose a line that fits the data well. In addition to this, there was particular interest in the

area surrounding the negative / positive cut-off point. It was essential to obtain a good fit there, since serosurvey samples with serological results around the positive / negative cut-off are the main candidates for misclassification.

Evaluating the model fit

Once the regression model was fitted, the strength of the line relationship or the proportion of data explained by the model was quantified by the *coefficient of determination* R^2 defined as:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2},$$

Equation 4.6

where \bar{y} is the mean of the data.

Based on data from the original ESEN project and the inherent variability of different assays, an R^2 of 0.80 or higher was considered to be “satisfactory” for the standardisation to take place. However, R^2 as low as 0.75 were also considered, especially for mumps or pertussis assays, for which there is higher variability due to weak correlates of protection (Andrews N *et al.*, 2000; Giammanco A *et al.*, 2003; Kafatos G *et al.*, 2005).

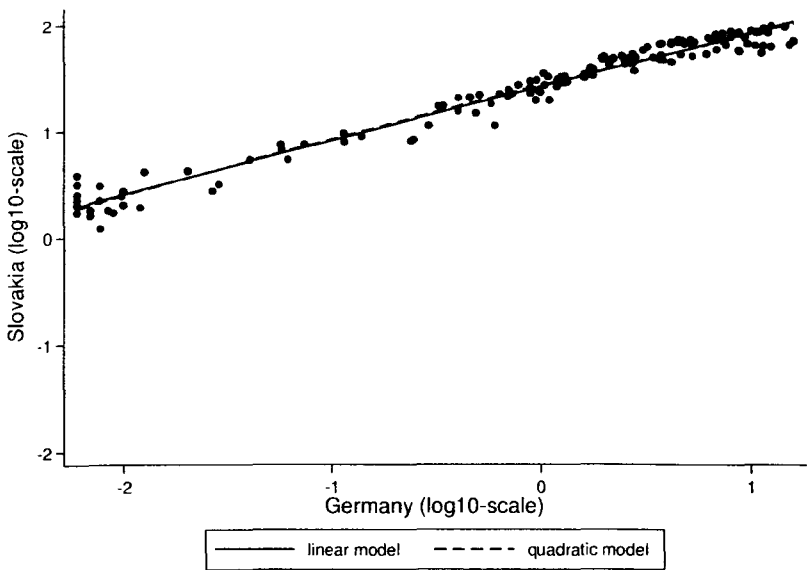
Type of regression model

A straight and a quadratic line equation were compared and, for parsimony, the simplest was sought i.e. simple linear regression was used unless there was significant evidence at the 5% level that the quadratic term was providing a better fit. In a few cases where there was a clear lack of fit around the critical area of the positive-negative cut-off, other models

were considered such as a cubic or sigmoid model (an example of a sigmoid model is given below).

Figure 4.4 shows an example of Slovakian measles panel results tested using a Virotech ELISA assay regressed against the German reference centre's results that used an Enzygnost ELISA assay. Both the linear and the quadratic curves are shown. Since the quadratic term was not significant, the simple linear regression model was chosen as the standardisation equation (linear regression t -test; $p = 0.16$).

Figure 4.4 Comparison between linear and quadratic curves for the standardisation panel tested against measles by the Slovakian and the German laboratories



In another example, the Czech Republic HAV panel is plotted against the Greek reference centre's results (Figure 4.5(a)). Unfortunately, the panel appears to have very few measurements around the assay cut-off of 0.01 IU/ml or -2 on the \log_{10} -scale where the fit matters most. This means it is quite difficult to estimate with precision the shape of the line around this area, and this could result in a high number of misclassifications when

applying the standardisation equation to the national serosurvey data. In this case the quadratic term is significant ($p < 0.001$). However, with the little information that is provided, it is clear that neither the linear nor the quadratic curves fit the data well around the crucial cut-off points. Hence, a sigmoid curve of the following type was proposed instead:

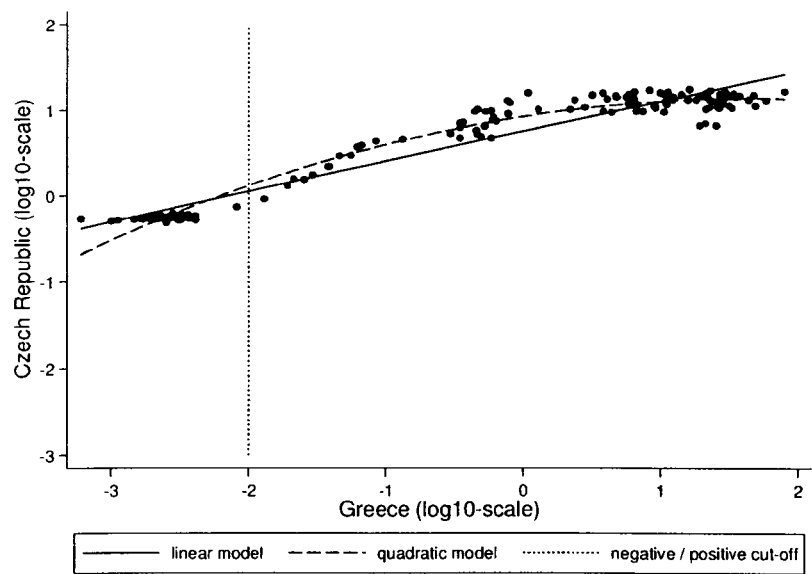
$$y_i = \alpha + \frac{\beta}{1 + e^{-(\gamma + \delta x_i)}} + \varepsilon_i ,$$

Equation 4.7

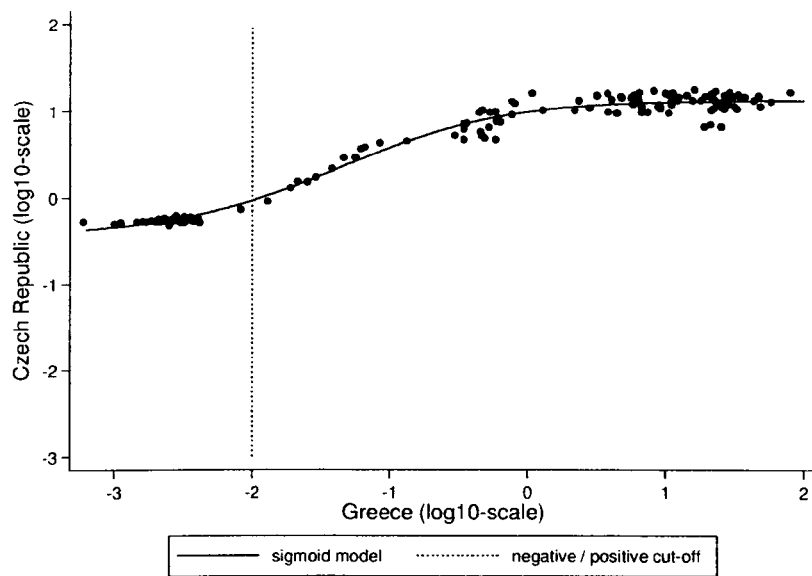
where $\alpha, \beta, \gamma, \delta \in \Re$ are constants and $\varepsilon_i \sim N(0, \sigma^2)$ (Figure 4.5(b)). Note that for the function $y(x) = \alpha + \frac{\beta}{1 + e^{-(\gamma + \delta x)}}$, the lower horizontal asymptote is $\lim_{x \rightarrow -\infty} (y(x)) = \alpha$ and the upper $\lim_{x \rightarrow +\infty} (y(x)) = \alpha + \beta$. Also the parameters γ and δ correspond to the location and slope parameters, respectively.

Figure 4.5 Comparison of different regression lines for the HAV panel tested by the Czech Republic and Greece laboratories

(a) Linear and quadratic curves



(b) Sigmoid regression line of the HAV panel for the Czech Republic plotted against the Greece reference centre's results



4.8 Effect of results outside the assay detection range

As mentioned earlier, many serological assays do not return quantitative results outside a specific range bound by detection limits. Although this has little effect for a patient's diagnosis, it can be a disadvantage for seroepidemiological studies, where all information is important. Such limited observations, where the only information available is whether their true value is below or above the detection range, are called *censored data*. Censored data are formally defined and discussed extensively in the next Chapter.

During the ESEN2 project, laboratories were encouraged to report quantitative results during the standardisation panel testing, however, in practice, this was not always possible. For analysis purposes, any results below the assay detection range (D_L, D_U) were substituted by $D_L / 2$ whereas any data above were replaced by $2D_U$ (Andrews N *et al.*, 2000; Giammanco A *et al.*, 2003; von Hunolstein C *et al.*, 2000).

In order to measure the effect of these conventions on the fitted line, two models were fitted, before and after excluding these observations. The model including results outside the detection range can be expressed as the function $\hat{y}(x_i)$, whereas the model that excludes these results as $\hat{y}_c(x_i)$. Given that the main interest is around the positive / negative cut-off point (say δ_0), the absolute difference was estimated between the two models on that point (i.e. $|\hat{y}(\delta_0) - \hat{y}_c(\delta_0)|$) where $\hat{y}(\delta_0)$ is defined as:

$$\hat{y}(\delta_0) = \hat{\beta}_0 + \hat{\beta}_1\delta_0 + \dots + \hat{\beta}_k\delta_0^k,$$

Equation 4.8

where $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$ are the parameter estimates of the model including censored data.

Similarly $\hat{y}_c(\delta_0)$ is defined as:

$$\hat{y}_c(\delta_0) = \hat{\beta}_{c0} + \hat{\beta}_{c1}\delta_0 + \dots + \hat{\beta}_{ck}\delta_0^k,$$

Equation 4.9

where $\hat{\beta}_{c0}, \hat{\beta}_{c1}, \dots, \hat{\beta}_{ck}$ are the regression parameter estimates after excluding any censored data.

To make the model difference $|\hat{y}(\delta_0) - \hat{y}_c(\delta_0)|$ comparable between different assays, it was divided by the range of the observations y_i , i.e. $\max(y_i) - \min(y_i)$. The resulting quantity, that will be referred to as *cut-off ratio (COR)*, measures the effect the censored data has on the regression line and is defined as follows:

$$COR = \frac{|\hat{y}(\delta_0) - \hat{y}_c(\delta_0)|}{\max(y_i) - \min(y_i)}.$$

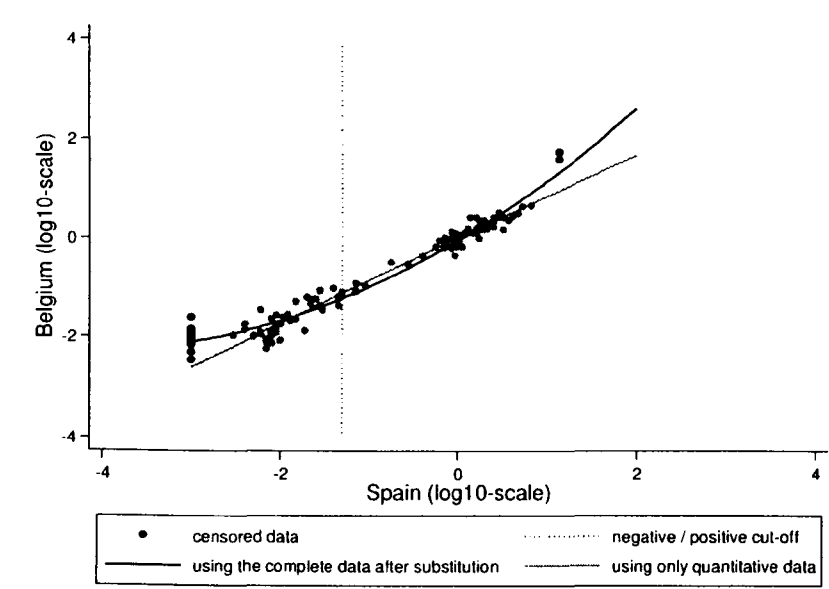
Equation 4.10

In the cases where the cut-off ratio was less than 0.05, there was little difference between the two regression lines and therefore the model based on the substituted censored observations was used. For $COR > 0.075$, the results outside the detection limit clearly influenced the model at the positive / negative cut-off, and hence, they were excluded from the analysis. Where the COR was between 0.05 and 0.075, the complete dataset was used unless the line excluding the censored data clearly improved the fit in the region of the positive / negative cut-off.

The impact of the results outside the detection range of the assay on the standardisation equation is shown in two examples. For the example given in Figure 4.6, the results of the

Belgian VZV panel testing for VZV were regressed against the Spanish reference centre's results. Spanish measurements below the detection range reported as "< 0.002 IU/ml" were substituted by 0.001, whereas those reported as "> 6.6 IU/ml" were substituted by 13.2. A quadratic equation was fitted twice, once using only the quantitative data, and once using all the data after substituting the censored observations. The difference between the two fitted lines on the cut-off point of 0.05 IU/ml was not considered important (*COR* = 0.04) and hence, the equation including all the observations was selected.

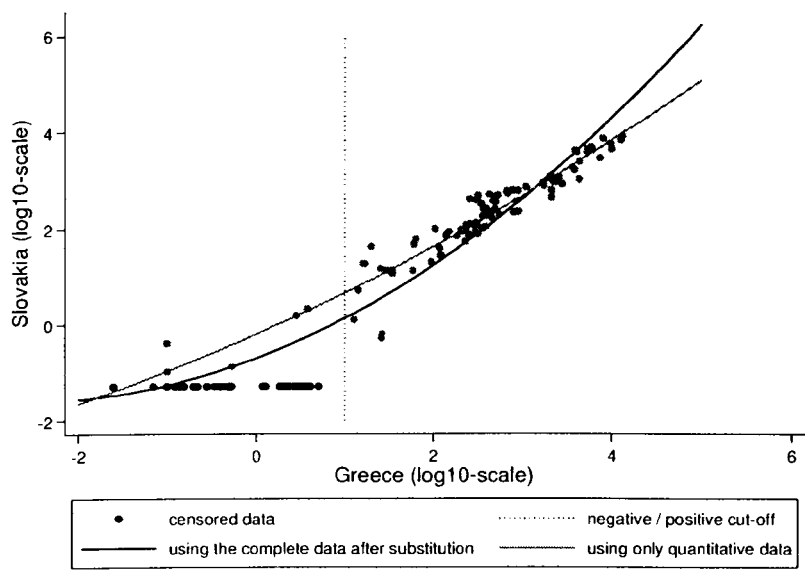
Figure 4.6 Influence of censored results on the regression line - VZV panel results for Belgium plotted against the Spanish reference centre's results



In another example, the results of the Slovakian panel testing for anti-HBs (MONOLISA antiHBs 3.0) were regressed against the Greek reference centre's results (AxSYM AUSAB assay). Seventy-five samples (45%) were below the assay detection range of the Slovakian assay and were reported as "< 0.1 units", whereas 15% of Greece assay were reported as "< 0.05". From Figure 4.7 it is evident that there are not enough quantitative results reported by Slovakia around the cut-off point. The censored data were substituted by

“0.05” and “0.025” for the Slovakian and the Greek assays, respectively. The difference between the equation using only the quantitative results and the one using all the data after substituting censored observations at the reference centre’s cut-off point (10 units), was $COR = 0.122$. Therefore, the equation based only on the quantitative result was selected as the standardisation equation (Figure 4.7).

Figure 4.7 Influence of censored results on the regression line - HBV panel results for Slovakia plotted against the Greece reference centre’s results



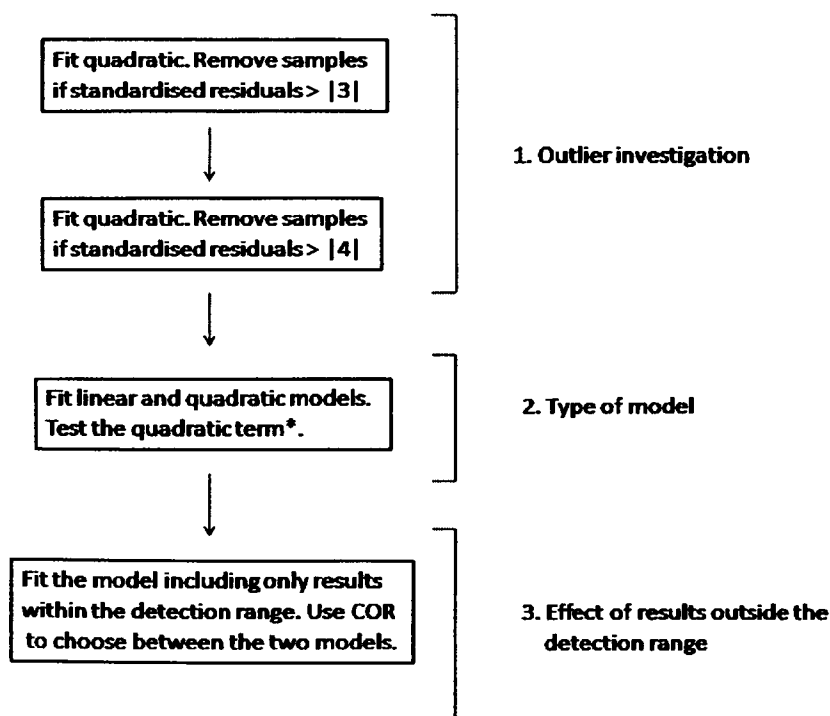
4.9 Standardisation algorithm

The methods used for dealing with outliers, selecting the type of equation and censoring measurements were all combined into one algorithm. The reason for this was that the methodology needed to be automated for the large quantities of data included in the ESEN2 project.

The flow-chart (Figure 4.8) presents a summary of the standardisation algorithm. Initially, potential outliers were flagged, investigated and either corrected or excluded from the next

steps of the analysis based on the more complex model (quadratic). The outlier investigation was carried first as any changes of the extreme points could affect the standardisation equation estimates. Linear and quadratic curves were then fitted and compared in order to select the type of the standardisation equation. Finally, the impact of censored data on the standardisation equation was assessed and if found to be influential, these data were also excluded.

Figure 4.8 Flow-chart summarising the standardisation algorithm



* Note that in exceptional situations where both the linear and the quadratic curves were clearly not appropriate an alternative model was preferred (e.g. cubic or sigmoid).

4.10 Back-standardisation

There were occasions when it was not possible to standardise a country’s results because the serosurvey was tested before the distribution of the panel. An alternative method of standardisation was then used named *back-standardisation*.

Following the serosurvey testing, approximately 150 samples were selected from the serosurvey after stratifying the results according to titre to ensure that the whole range from high positive to low negative were represented. The samples were sent to the reference laboratory to be re-tested. The standardisation was then performed in the same way as for the standardisation panel.

4.11 Standardised vs. non-standardised seroprevalence estimates

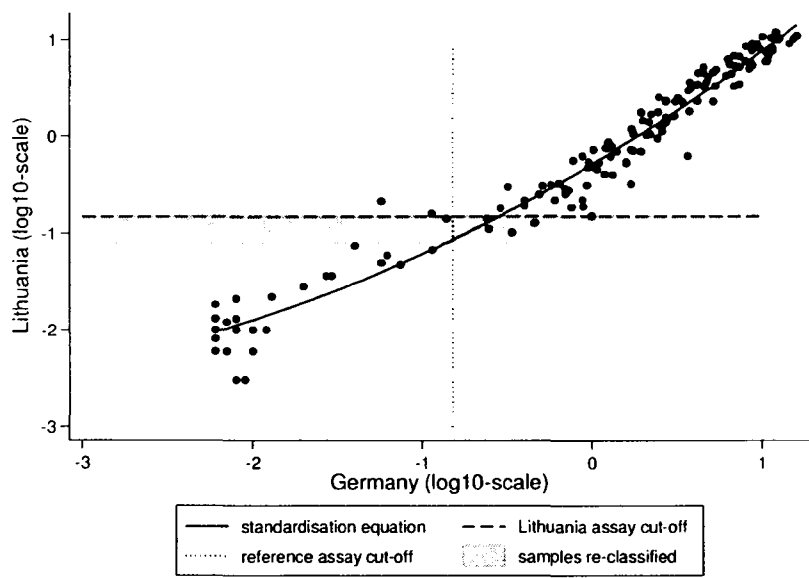
To demonstrate the impact of the standardisation equation on estimating seroprevalence the previous example of the measles panel tested by Lithuania and Germany is revisited. After applying the standardisation algorithm, a fitted quadratic equation of the same type as shown in Equation 4.4 was chosen to standardise the serosurvey outcome:

$$\hat{y}_i = -0.29 + 1.05x_i + 0.12x_i^2,$$

where $i = 1, \dots, 148$, x_i are the \log_{10} -transformed measurements given by the German laboratory and \hat{y}_i are the predicted \log_{10} -transformed Lithuanian results.

Figure 4.9 shows that a quadratic relationship gives a satisfactory description of the data. However, it is clear that any serosurvey samples classified as “negative” by the Lithuanian assay but measured close to the Lithuanian positive / negative cut-off of 0.15 IU/ml (shaded area) could be re-classified as “positive” by the reference assay following the application of the standardisation algorithm.

Figure 4.9 Standardisation curve for Lithuania measles panel results



Following the serosurvey testing, the testing laboratory results can be transformed into the same scale as the reference laboratory’s results. For a quadratic equation, a testing laboratory measurement y_0 can be transformed to reference centre units x_0 as follows:

$$x_0 = \frac{-\hat{\beta}_1 + \sqrt{\hat{\beta}_1^2 - 4\hat{\beta}_0(\hat{\beta}_2 - y_0)}}{2\hat{\beta}_0},$$

Equation 4.11

where $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are the regression parameters estimated from Equation 4.4.

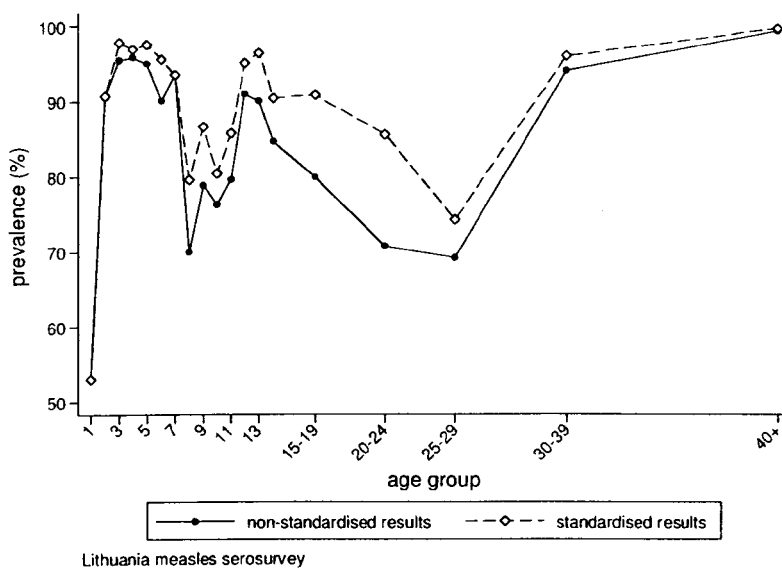
Using Equation 4.11, a Lithuanian measurement y_0 was transformed into the same unit scale as the reference laboratory’s measurement x_0 as follows:

$$x_0 = \frac{-1.05 + \sqrt{1.05^2 - 4 \times 0.12 \times (-0.29 - y_0)}}{2 \times 0.12}.$$

As an example, a measurement of 0.1 IU/ml (or -1 on the \log_{10} -scale) which was below the 0.15 IU/ml cut-off by the Lithuanian assay and was initially classified as negative, would be transformed to 0.18 IU/ml (or -0.73 on the \log_{10} -scale), and therefore, reclassified as positive.

Figure 4.10 shows the seroprevalence estimates before and after standardisation. A number of samples that have been re-classified have been taken from individuals aged between 3 and 6 years and also 8 and 29 years. The high number of borderline measurements for these age groups may reflect the lower antibody response and waning immunity with time following vaccination.

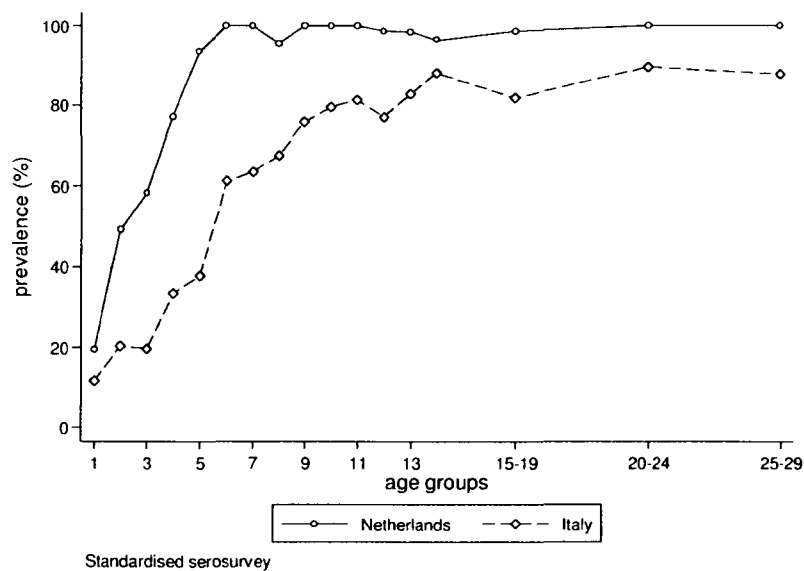
Figure 4.10 Seroprevalence estimates for Lithuania measles results



As a second example, the Italian and Dutch serosurveys presented in Section 4.2 are revisited. As it shown in Figure 4.11, there is little difference in seroprevalence estimates

following standardisation. This means that seroprevalence differences can be attributed to genuine seroepidemiological disparities instead of difference between laboratory methods.

Figure 4.11 Age-specific seroprevalence estimates of VZV antibody for the Netherlands and Italy (standardised)



4.12 Conclusions

This Chapter presented a collection of methods developed to generate equations for standardising serological outcomes tested in different laboratories into comparable units. Separate techniques for dealing with issues such as selecting the type of equation, outliers and measurements outside the assay detection limits, were brought together into one automated algorithm to be used for the ESEN2 project (Kafatos G *et al.*, 2005). Once the serological outcome was standardised using this automated routine, comparable seroprevalence estimates were produced.

There are a number of reasons as to why seroprevalence estimates differ following standardisation and sample re-classification. This can be due to variability of the

serological measurements occurring when samples are tested by different assays or laboratories. In addition, higher variability can be expected from assay tests of specific antigens which may be due to cross-reactivity as explained in Section 2.8. An attempt to quantify the impact standardised results have on seroprevalence estimation and to examine the impact of the factors above will be presented in Chapter 6.

Chapter 5: Censored Serological Data

5.1 Assay detection limits

As discussed in the previous Chapter, one important issue that affects standardisation equations is that many serological assays are constrained by *detection limits (DL)*. One reason for setting a detection range could be the failure of instruments to detect levels below or above certain values (say D_L and D_U , respectively). A lower threshold is common for immunoassays such as the ELISA that require antigen concentrations sufficient for binding antibodies (Schisterman EF *et al.*, 2006). Low concentrations may not be reliably differentiated from the background noise, and hence, not accurately measured. An upper detection limit may occur in dilution assays in cases of extremely high antibody response, when testing for antibodies at an additional dilution is not considered worthwhile (Barnhart HX *et al.*, 2005; Schisterman EF *et al.*, 2006; Whitcomb BW & Schisterman EF, 2008).

For the ESEN2 project, during the testing of the standardisation panel, laboratories were asked to provide quantitative measurements even outside the assay range. For data below D_L , this sometimes meant having higher measurement error due to the background noise (Whitcomb BW & Schisterman EF, 2008; Lim J, 2006). However, it is always good practice to obtain such measurements, since the thresholds assigned to distinguish between values with and without background noise may not be appropriate for seroepidemiological studies (Whitcomb BW & Schisterman EF, 2008). As mentioned previously, many laboratories were not able to produce quantitative measurements throughout the assay range.

The analysis of biological assays constrained by thresholds is an area that has been developing rapidly during the last five years, with numerous peer-reviewed papers and techniques published. The aim of this Chapter to review some of these methods and to use them to generate standardisation equation estimates that are comparable to the ones obtained by the ESEN2 methods.

5.2 Types of censored data

In statistical terms, a set of data is called *censored* when observations are known only to be greater (or less) than some value (Bland M, 2000). This is the case with serological data when quantitative measurements (the exact values of which may be known or unknown) outside the assay DL are reported as fixed values. A distinction needs to be made between censored and *truncated* data, the latter occurring when values outside a certain range are entirely omitted (Breen R, 1996;Greene WH, 2003;Maddala GS, 2001).

There are different types of censored data, however, the following three categories are considered here:

- (i) *Right-censored*, where all that is known is that the data are higher than a specific value. In terms of serological assays, these are the measurements reported as “ $> D_U$ ” (Zhang Z & Sun J, 2010).
- (ii) *Left-censored*, data where all that is known is that the data are less than a specific value. In terms of serological assays, these are the measurements reported as “ $< D_L$ ”. Although assays can be bounded from both sides, left-censoring appears more commonly, and therefore, the analysis will be mainly concentrating on these scenarios (Baccarelli A *et al.*, 2005;Barnhart HX *et al.*, 2005;Zhang Z & Sun J, 2010).
- (iii) There is a third type of censored data which includes measurements where all that is known is that they lie within an interval (Zhang Z & Sun J, 2010). This type of censoring will be discussed in more detail later in this Chapter.

5.3 Deletion and simple substitution methods

There are a number of different methods that have been proposed for dealing with censored biological assay data. The simplest is the *deletion* method that omits the censored observations from the analysis resulting in a truncated dataset. This has obvious disadvantages given there is loss of information (Baccarelli A *et al.*, 2005; Jain RB *et al.*, 2008; Schisterman EF *et al.*, 2006).

Another easy-to-apply method involves analysing the dataset after substituting censored observations by a single value. This method is called *simple substitution*. Various recipes have been proposed for estimating the substituted value. For left-censored data, they include substituting censored data with 0, D_L , $D_L / 2$, $D_L / \sqrt{2}$ or $2D_L / 3$ (Baccarelli A *et al.*, 2005; Barnhart HX *et al.*, 2005; Daniels RD & Yiin JH, 2006; Jain RB *et al.*, 2008; Krishnamoorthy K *et al.*, 2009; Richardson DB & Ciampi A, 2003; Schisterman EF *et al.*, 2006; Lubin JH *et al.*, 2004). Substitution with 0 (which is not applicable to the log-transformed data presented here) has been found to overestimate the slope when fitting a straight line. On the other hand, replacing with D_L seems to produce the opposite effect i.e. underestimation of slope (Helsel DR, 2005; Lubin JH *et al.*, 2004). Substitution with $D_L / \sqrt{2}$ is appropriate if the data follow a normal distribution (e.g. log-transformed serological data), and $D_L / 2$ can be applied for highly-skewed data (e.g. non-transformed serological data) (Jain RB *et al.*, 2008). These substitutions have been found to work and are particularly useful for low proportions of censored data.

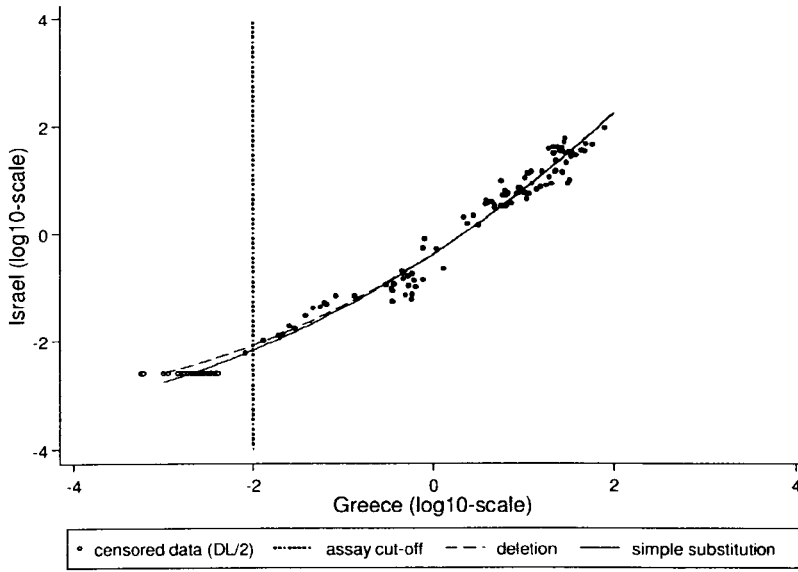
As mentioned in the previous Chapter, standardisation equations were estimated by both deletion and simple substitution methods ($D_L / 2$ and $2D_U$ for left- and right-censoring, respectively) prior to log-transformation. To recapitulate, given the importance of good fit around the positive / negative cut-off point, the estimated equations by the two methods

were compared at that point. The simple substitution method was generally used, apart from situations where the difference between the methods at the point of interest was large and the deletion method was preferred (Chapter 4) (Kafatos G *et al.*, 2005).

An example of left-censoring from the ESEN2 project is as follows: Israel's HAV panel results, reported as IU/ml, were plotted after they were log-transformed against the Greece reference centre's results that had a positive / negative cut-off of 0.01 IU/ml (-2 on the \log_{10} -scale). Of the 142 samples tested by both laboratories after excluding the outliers, 39 (27%) were reported as "< 0.005" IU/ml by Israel.

Using the ESEN2 standardisation methodology, two quadratic regression equations were produced using deletion and simple substitution by $D_L / 2$, after assuming the data followed a log-normal distribution (i.e. replacing the censored observations by $\log_{10}(0.005 / 2)$). For this example, simple substitution was chosen for the appropriate standardisation equation instead of deletion, since the difference between the two lines at the reference centre's positive / negative cut-off was not large ($COR = 0.019$, using Equation 4.10). Although for higher measurements the two fitted lines gave similar fitted values, this was not the case for lower measurements, where the model using simple substitution was affected by the substituted observations (Figure 5.1).

Figure 5.1 Regression curves using deletion and simple substitution



A comparison of standardisation equation estimates for the deletion and the simple substitution methods was carried out. The cut-off values for Israel ($\hat{y}_c(-2)$ and $\hat{y}(-2)$ for deletion and simple substitution, respectively) were estimated, after substituting the reference assay cut-off of -2 (on the \log_{10} -scale), as in Equations 4.8 and 4.9:

$$\hat{y}(-2) = \hat{\beta}_0 + \hat{\beta}_1(-2) + \hat{\beta}_2(-2)^2 = -0.39 + 1.10(-2) + 0.10(-2)^2 = -2.18$$

and

$$\hat{y}_c(-2) = \hat{\beta}_{c0} + \hat{\beta}_{c1}(-2) + \hat{\beta}_{c2}(-2)^2 = -0.38 + 1.08(-2) + 0.11(-2)^2 = -2.08.$$

where $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$ were the parameter estimates for the simple substitution and $\hat{\beta}_{c0}$, $\hat{\beta}_{c1}$

$\hat{\beta}_{c2}$ for the deletion method.

Table 5.1 shows a comparison between the regression and cut-off estimates using the different methods. The standardised cut-off was estimated highest after substituting by D_L ($\hat{y}(-2) = -2.01$). This is expected, since the censored observations, due to their high value, pull the lower end of the equation upwards towards 0. The standardised cut-off was estimated lowest after substituting by $D_L/2$ ($\hat{y}(-2) = -2.18$). This is due to the low values of the data used for substitution that pull the lower end of the equation downwards.

Table 5.1 Standardisation equation estimates using simple substitution and deletion methods (95 % CIs in brackets)

Method		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	cut-off
Deletion		-0.38 (-0.45, -0.31)	1.08 (1.04, 1.13)	0.11 (0.07, 0.16)	-2.08
simple	D_L	-0.41 (-0.46, -0.35)	1.08 (1.04, 1.11)	0.14 (0.12, 0.16)	-2.01
Substitution	$D_L / 2$	-0.39 (-0.45, -0.33)	1.10 (1.07, 1.14)	0.10 (0.08, 0.12)	-2.18
	$D_L / \sqrt{2}$	-0.40 (-0.45, -0.34)	1.09 (1.06, 1.12)	0.12 (0.10, 0.14)	-2.10
	$2D_L / 3$	-0.40 (-0.45, -0.34)	1.09 (1.06, 1.13)	0.12 (0.10, 0.14)	-2.11

5.4 Simulations for regression equations

The different methods were validated using a simulation technique similar to a parametric *bootstrap* (Efron B & Tibshirani RJ, 1993). Censored data were generated randomly from the reference centre’s results using a model of the general form shown in Equation 4.1 where $\beta_0, \beta_1, \dots, \beta_k$ and σ were regarded as known since they had been chosen in advance. Such models to be used for simulation-generating scenarios will be referred to as *true models*.

The process is described as follows:

1. Based on the number of samples included in the reference panel (say n), a set of random data was generated from a distribution with mean $\mu = 0$ and variance σ^2 , i.e. $N(0, \sigma^2)$, to obtain the residuals.

2. Assuming the generated data represent a set of residuals $\hat{\varepsilon}_i^*$ where $i = 1, \dots, n$, a set of measurements for the testing laboratory y_i^* (the laboratory that is regressed against the reference laboratory) was generated using the equation

$$y_i^* = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k + \hat{\varepsilon}_i^* .$$

Equation 5.1

3. A detection range (D_L, D_U) was applied to the simulated data by assuming that any observations below the range were replaced by “ $< D_L$ ” and any observation above by “ $> D_U$ ”.

4. For the simulated dataset, regression equations were fitted, producing a set of estimates for each of the methods described above i.e. $\hat{\beta}_{C0}^*, \hat{\beta}_{C1}^*, \dots, \hat{\beta}_{Ck}^*$ where C denotes the method used (deletion or any of the simple substitution methods - Note the change of notation up to this point since the subscript C denoted the deletion method only). Together with the regression estimates, the cut-off value for the testing laboratory $\hat{y}_C^*(\delta_0)$, where δ_0 denotes the reference assay cut-off, was estimated using Equations 4.8 and 4.9.

5. The process was repeated multiple times (say M) and the mean of the estimates for each parameter was calculated, i.e. $\frac{1}{M} \sum_{j=1}^M \hat{\beta}_{c0j}^*$, $\frac{1}{M} \sum_{j=1}^M \hat{\beta}_{c1j}^*$, ..., $\frac{1}{M} \sum_{j=1}^M \hat{\beta}_{ckj}^*$, as well as the cut-off point $\frac{1}{M} \sum_{j=1}^M y_c^*(\delta_0)$.

The amount of uncertainty in estimates arising from the variability between the simulated datasets is quantified by two different methods:

(a) The *percentile interval*. This includes the 2.5th and 97.5th percentiles of the simulated estimates as the lower and upper limits, respectively.

(b) The *coverage probability*. This is the proportion of simulation runs for which the confidence interval for each of the estimated parameters $\hat{\beta}_{c0}^*, \hat{\beta}_{c1}^*, \dots, \hat{\beta}_{ck}^*$ at that run contains the true model parameters $\beta_0, \beta_1, \dots, \beta_k$.

The performance of a method C is then evaluated by comparing the mean of the estimates of the parameters to their true values, by checking whether the true values lie in the percentile intervals and whether the coverage probability is close to its nominal value.

5.5 Comparison of deletion and simple substitution methods using simulations

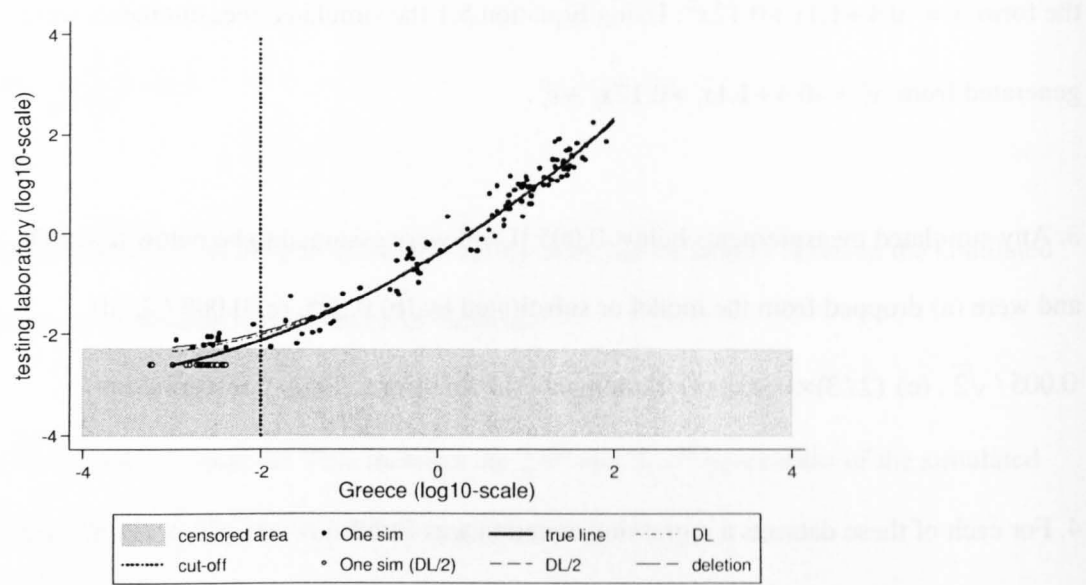
The simulation process described above was applied to the Israel HAV panel example comparing the deletion with a number of simple substitution methods.

1. Based on the number of samples included in the Greek reference panel, 148 random samples were generated from a normal distribution model with mean $\mu = 0$ and standard deviation $\sigma = 0.26$ to be used as residuals $\hat{\varepsilon}_i^*$.

2. The measurements of the testing laboratory were then generated from a true model of the form $y = -0.4 + 1.1x + 0.12x^2$. Using Equation 5.1 the simulated measurements were generated from $y_i^* = -0.4 + 1.1x_i + 0.12x_i^2 + \hat{\varepsilon}_i^*$.
3. Any simulated measurements below 0.005 IU/ml were assumed to be below assay DL and were (a) dropped from the model or substituted by (b) 0.005, (c) $0.005 / 2$, (d) $0.005 / \sqrt{2}$, (e) $(2/3) \times 0.005$ (measurements shown prior to \log_{10} -transformation).
4. For each of these datasets a regression equation was fitted.
5. The data were simulated 1,000 times and the average of the estimates was taken. The coverage probability was chosen to describe the uncertainty around the regression parameters, whereas the 95% percentile interval was selected for the cut-off estimate.

Figure 5.2 shows the model comparison between deletion and substitution by 0.005 and $0.005 / 2$. Data from one set of simulations is shown as an example. The regression equation after substituting with $0.005 / 2$ gave estimates closer to the truth for antibody measurements around the cut-off of interest (-2 on the \log_{10} -scale), and lower when compared to the other two methods.

Figure 5.2 Comparison between simple substitution and deletion methods after 1,000 simulations generated from $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$



Note: for graphical purposes one set of simulated results was plotted

The means of the parameter estimates with their corresponding coverage probabilities for the 95% confidence intervals are shown in Table 5.2. The positive / negative cut-off is also presented together with its corresponding 95% percentile interval. For the true model $y = -0.4 + 1.1x + 0.12x^2$, the cut-off was $y(-2) = 2.12$ (Equations 4.8 and 4.9). Simple substitution with $D_L / 2$ gave coverage probabilities close to the nominal value and a cut-off estimate very close to the truth ($\hat{y}_C^*(-2) = -2.13$). The other two methods with high coverage were substitution with $D_L / \sqrt{2}$ and $2D_L / 3$.

Table 5.2 Mean parameter estimates and 95% coverage probabilities generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$

Method		$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	cut-off*
true model		-0.40	1.10	0.12	-2.12
Deletion		-0.42 (90.6%)	1.08 (81.2%)	0.15 (39.6%)	-1.96 (-2.03, -1.89)
simple	D_L	-0.43 (83.5%)	1.09 (88.1%)	0.15 (26.7%)	-2.00 (-2.06, -1.96)
Substitution	$D_L / 2$	-0.41 (93.2%)	1.10 (93.0%)	0.12 (94.8%)	-2.13 (-2.19, -2.06)
	$D_L / \sqrt{2}$	-0.42 (89.0%)	1.10 (92.7%)	0.14 (69.3%)	-2.07 (-2.13, -2.01)
	$2D_L / 3$	-0.42 (89.9%)	1.10 (92.3%)	0.13 (76.2%)	-2.08 (-2.14, -2.02)

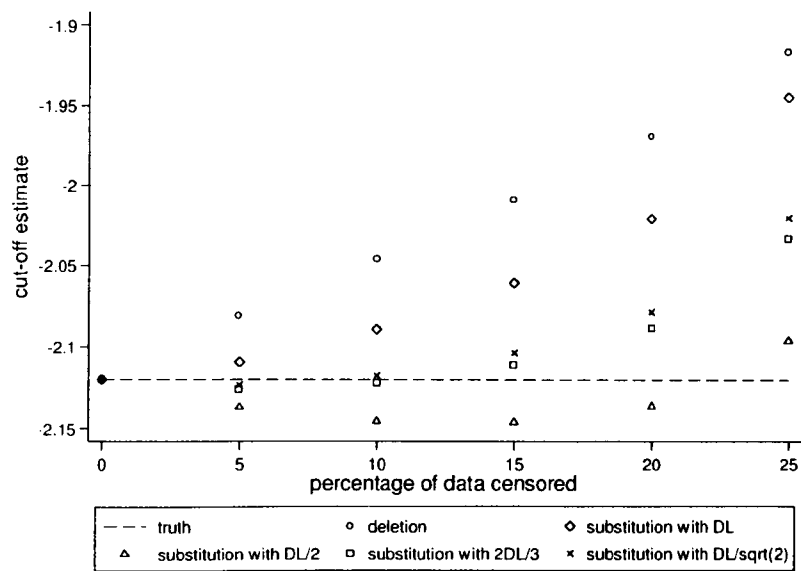
* 2.5th and 97.5th percentiles given instead of coverage probabilities

The example shown above and the generated simulations examine the scenario where the Israel panel results were reported as censored if below 0.005 IU/ml. This resulted in 27% of censored data. However, the methods presented are expected to vary in performance according to the proportions of data being censored. Simulations were thus performed in the same way as explained above, varying the point where the data were censored.

Figure 5.3 shows a comparison of the cut-off estimates between the different methods. For up to 15% of data censored substitution with $2D_L / 3$ and $D_L / \sqrt{2}$ produced cut-off estimates closest to the truth compared to other methods, followed by substitution with $D_L / 2$. For 20% to 25% of data censored, substitution with $D_L / 2$ produced the best estimates.

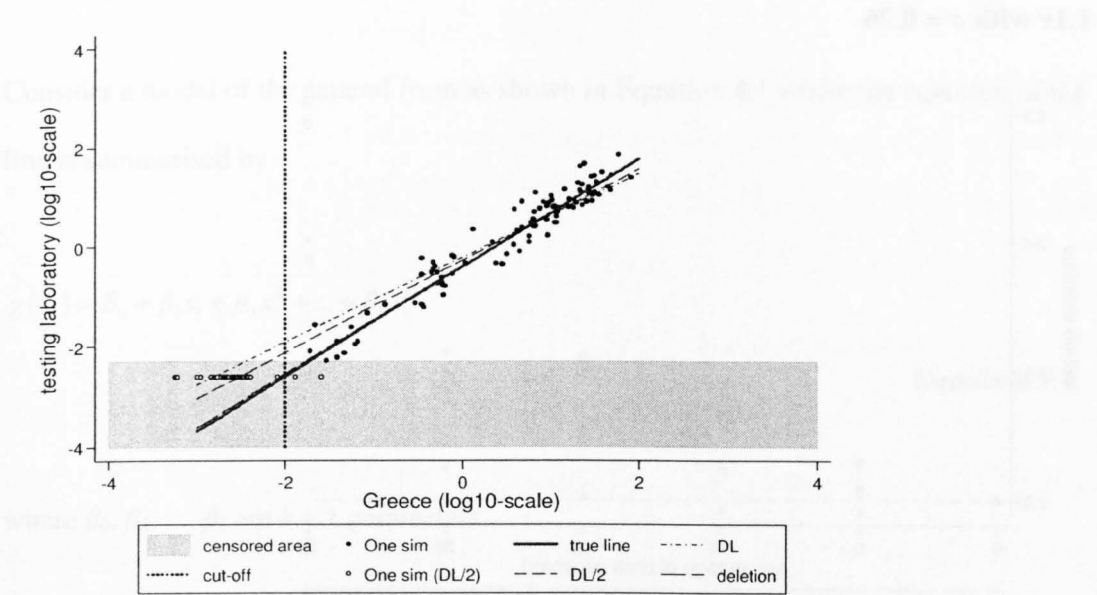
A maximum of 25% of the data were censored in the simulated datasets, given that it would make no sense to have censored observations above the standardised cut-off value of -2.12 (from Table 5.2).

Figure 5.3 Comparison cut-off estimate between simple substitution and deletion methods for different proportions of data censored generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$



The method comparison was repeated after generating simulations from a simple linear regression model of the form $y = -0.4 + 1.1x$ with $\sigma = 0.26$. Assuming censoring of data below 0.005 IU/ml, the model using deletion gave better estimates than the simple substitution methods (Figure 5.4). According to the standardisation algorithm for such a scenario, the regression based on deletion would have been selected to obtain the standardisation equation, due to the large differences between this method and the simple substitution regression on the cut-off point (Chapter 4). The large differences between the two methods may be partly due to the fact that the 0.005 threshold value is close to the assay cut-off.

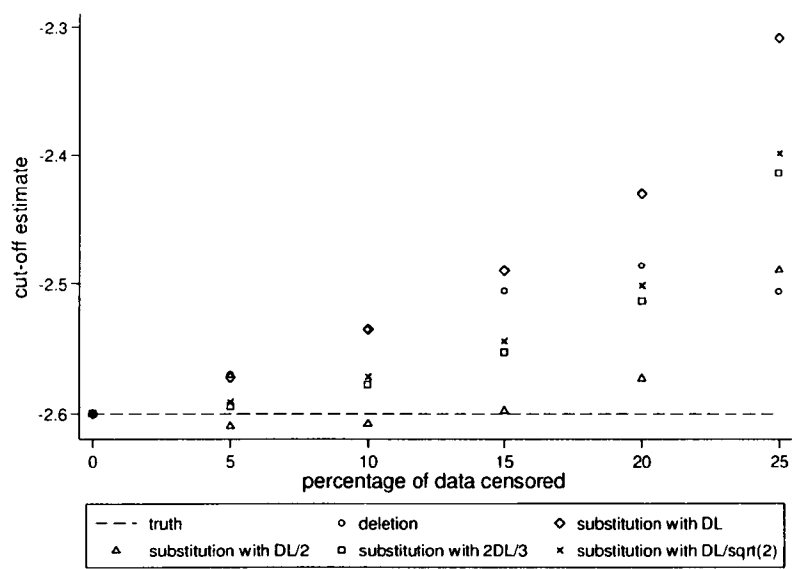
Figure 5.4 Comparison between simple substitution and deletion methods after 1,000 simulations generated from $y = -0.4 + 1.1x$ with $\sigma = 0.26$



Note: for graphical purposes one set of simulated results was plotted

For the true model $y = -0.4 + 1.1x$, the cut-off was $y(-2) = -2.6$. For up to 20% of the data being censored, substitution with $D_L / 2$ gave a consistently better estimate at the point of cut-off in comparison to the other methods. However, for 25% of the data being censored, all methods produced biased results, with the deletion method having the closest estimate to the truth (Figure 5.5).

Figure 5.5 Comparison of cut-off estimate between simple substitution and deletion methods for different proportions of data censored generated from the model $y = -0.4 + 1.1x$ with $\sigma = 0.26$



5.6 Censored regression

Censored regression analysis is a statistical tool for incorporating observations limited by censoring on the response variable. It is a generalisation of the normal regression, and if the number of censored data is small, then censored regression tends to give similar estimates to normal regression. Censored regression was initially introduced as an extension of the probit model for an econometrics application by Tobin, and was later referred to as “Tobit regression” (Lubin JH *et al.*, 2004). Several types of censored regression models exist, however, since the Tobit model is the simplest and most commonly used amongst them, it will be used throughout hereafter, and will simply be referred to as censored regression. There are different ways of estimating the parameters of such models using *ordinary least squares (OLS)* and *maximum likelihood (ML)* techniques (Breen R, 1996;Jain RB *et al.*, 2008). The latter will be used in this Chapter, since the

estimators with ML techniques appear to have some advantages when compared to the OLS method (Breen R, 1996).

Consider a model of the general form as shown in Equation 4.1 where the equation of the line is summarised by:

$$g(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \dots + \beta_k x_i^k ,$$

Equation 5.2

where $\beta_0, \beta_1, \dots, \beta_k$ are $k + 1$ parameters.

Suppose the observed response variable y_i consists of (a) n measurements of which the first ℓ are quantitative, (b) a number of left-censored observations $y_i = D_{Li}$ for $i = \ell + 1, \dots, c$ where all that it is known is that the true value $y_i < D_{Li}$ and (c) a number of right-censored observations $y_i = D_{Ui}$ for $i = c + 1, \dots, n$ where all it is known is the true value $y_i > D_{Ui}$. Note that the subscript i is added to the notation D_L and D_U to account for potentially different censoring bounds. Let also $f(y; g(x_i), \sigma)$ be the *probability density function (PDF)* and $F(y; g(x_i), \sigma)$ the *cumulative distribution function (CDF)* with equation of the line $g(x_i)$ and standard deviation σ . The likelihood for the censored regression model is:

$$L = \prod_{i=1}^{\ell} f(y_i; g(x_i), \sigma) \times \prod_{i=\ell+1}^c F(y_i; g(x_i), \sigma) \times \prod_{i=c+1}^n (1 - F(y_i; g(x_i), \sigma)).$$

Equation 5.3

The log-likelihood is then:

$$\ln(L) = \sum_{i=1}^{\ell} \ln \left(f \left(y_i; g(x_i), \sigma \right) \right) + \sum_{i=\ell+1}^c \ln \left(F \left(y_i; g(x_i), \sigma \right) \right) + \sum_{i=c+1}^n \ln \left(1 - F \left(y_i; g(x_i), \sigma \right) \right).$$

Equation 5.4

The parameters included in equation $g(x_i)$ and the standard deviation σ can then be estimated by using maximum likelihood estimation (Lubin JH *et al.*, 2004).

5.7 Method of Multiple Imputation (MI)

Apart from deletion, simple substitution methods and censored regression, other methods have been proposed in the literature for dealing with censored observations. A method often quoted is Multiple Imputation (MI). The main reason for using MI would be to obtain explicit values for the censored data that can be used in further analysis. Given that the main aim is to estimate regression parameters, the censored regression method discussed above should be sufficient (Lubin JH *et al.*, 2004). However, given that a variety of MI methods are being used in the literature to obtain regression estimates from censored data, it seems necessary to compare it to the other methods. A simple application of the MI method consists of the following steps:

1. For a model of the type defined in Equation 4.1, use censored regression to obtain initial parameter estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ and $\hat{\sigma}$.
2. A set of residuals $\hat{\varepsilon}_i^*$ can then be generated from $N(0, \hat{\sigma}^2)$ using Equation 5.1.
3. After restricting the data to the censored observations $i = \ell + 1, \dots, n$ (including left- as well as right-censored data), a set of imputed quantitative measurements \hat{y}_i^* can be obtained for the censored data.

4. Using the original dataset with the addition of the imputed observations obtained from Step 3, a regression model of the general form shown in Equation 4.1 can be fitted and MI estimates $\hat{\beta}_0^*, \hat{\beta}_1^*, \dots, \hat{\beta}_k^*$ and $\hat{\sigma}^*$ can be achieved.

5. The procedure can be repeated multiple times resulting in estimates $\hat{\beta}_{0j}^*, \hat{\beta}_{1j}^*, \dots, \hat{\beta}_{kj}^*$ and $\hat{\sigma}_j^*$ where $j = 1, \dots, m$ denotes the number of imputations. The means of these estimates i.e.

$\frac{1}{m} \sum_{j=1}^m \hat{\beta}_{0j}^*, \frac{1}{m} \sum_{j=1}^m \hat{\beta}_{1j}^*, \dots, \frac{1}{m} \sum_{j=1}^m \hat{\beta}_{kj}^*$ are the MI estimates. The number of imputations does not

need to be large, with a recommended value between 3 and 5, and will need to be larger only if a larger proportion of the data is missing. For the examples shown below, $m = 10$ was used to fully account for the imputation variance, as has been done in previous studies (Lubin JH *et al.*, 2004).

As with censored regression, several variations of the MI method have been proposed in the past. One of these methods, proposed by Lubin, includes bootstrapping for obtaining the initial parameter estimates $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k$ and $\hat{\sigma}$ to be used for the imputations (Lubin JH *et al.*, 2004). Other methods include *Markov Chain Monte Carlo (MCMC)* adapted for MI (Schafer JL, 2010). The procedure explained above, a very basic application of MI, is the one that will be examined throughout this Chapter.

5.8 Confidence intervals for MI estimates

By combining the two variance components, the within- and between-imputation variance, it is possible to obtain a pooled variance and hence, confidence intervals around the MI estimates (Krishnamoorthy K *et al.*, 2009).

Suppose we are interested in calculating the variance of the MI estimates:

$$\bar{\beta}_K^* = \frac{1}{m} \sum_{j=1}^m \hat{\beta}_{Kj}^* ,$$

Equation 5.5

where $K = 0, 1, \dots, k$ denotes the number of parameters and m the number of imputations.

The within-imputation variance is defined by:

$$WI = \frac{1}{m} \sum_{j=1}^m \hat{\sigma}_j^{*2} .$$

Equation 5.6

The between-imputation variance can be defined as:

$$BI_K = \frac{1}{m-1} \sum_{j=1}^m (\hat{\beta}_{Kj}^* - \bar{\beta}_K^*)^2 .$$

Equation 5.7

The pooled variance for each estimate is given by combining the within-imputation variance in Equation 5.6 and the between-imputation variance in Equation 5.7 as follows:

$$T_K = WI + \left(1 + \frac{1}{m}\right) BI_K ,$$

Equation 5.8

where the term $\left(1 + \frac{1}{m}\right)$ compensates for the finite number of imputations.

Confidence intervals may be obtained from:

$$\bar{\beta}_K^* \pm t_{df_K, \alpha/2} \sqrt{T_K},$$

Equation 5.9

where $t_{df_K, \alpha/2}$ denotes the $(1 - \alpha)$ quantile of a Student's t-distribution with degrees of freedom:

$$df_K = (m - 1) \left(1 + \frac{mWI}{(m + 1)BI_K} \right)^2$$

Equation 5.10

(Carpenter J & Goldstein H, 2004; Schafer JL, 2010).

5.9 Example and simulations

Using the same example as above, the HAV Israel panel results were regressed against data from the Greek reference centre using a censored regression model by maximising the log-likelihood in Equation 5.4:

$$\ln(L) = \sum_{i=1}^{103} \ln(f(y_i; g(x_i), \sigma)) + \sum_{i=104}^{142} \ln(F(y_i; g(x_i), \sigma)),$$

where $g(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2$ for $i = 1, \dots, 142$ observations of which 39 were left-censored and $f(\cdot)$ and $F(\cdot)$ are the normal PDF and CDF, respectively.

For obtaining the MI estimates the following steps were implemented:

1. The estimates $\hat{\beta}_0 = -0.33$, $\hat{\beta}_1 = 1.13$, $\hat{\beta}_2 = 0.04$, $\hat{\sigma} = 0.24$ were obtained from censored regression.

2. A set of residuals $\hat{\varepsilon}_i^*$ was generated from a normal distribution $N(0, 0.24^2)$.

3. For the censored observations included in the sample (i.e. 39 samples were reported as “< 0.005” IU/ml) a set of imputed measurements was generated using

$$y_i^* = -0.33 + 1.13x_i + 0.04x_i^2 + \hat{\varepsilon}_i^*.$$

4. A quadratic regression model was fitted on the imputed dataset.

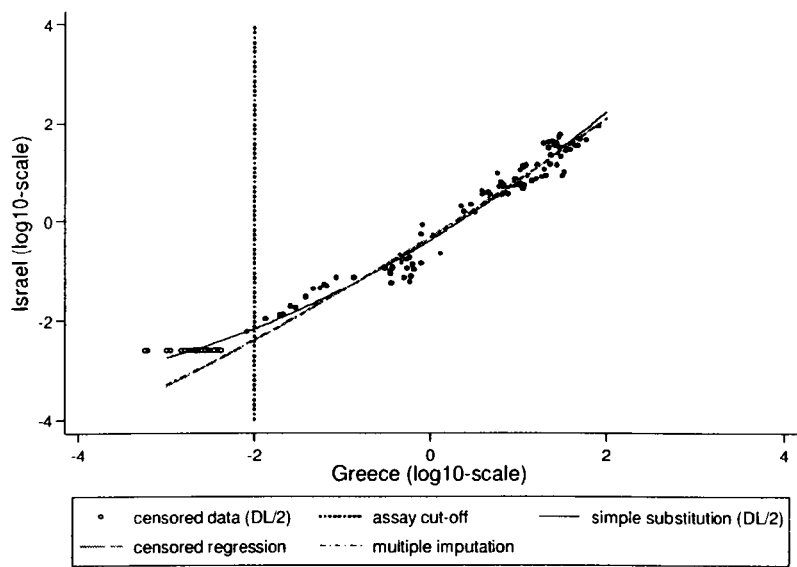
5. The process was repeated 10 times and the average of each estimate was taken as the MI estimate.

Confidence intervals were calculated around the MI estimates after combining the within- and between-imputation variance as shown in Section 5.8.

The resulting regression equations using the different methods are depicted in Figure 5.6.

The censored regression method was compared to the MI method and the simple substitution using $D_L/2$. In this example, the censored regression and MI gave regression equations similar to each other, whereas the regression using simple substitution method was highly influenced by the censored observations at the lower end.

Figure 5.6 Regression curves for Israel HAV panel using simple substitution, censored regression and MI



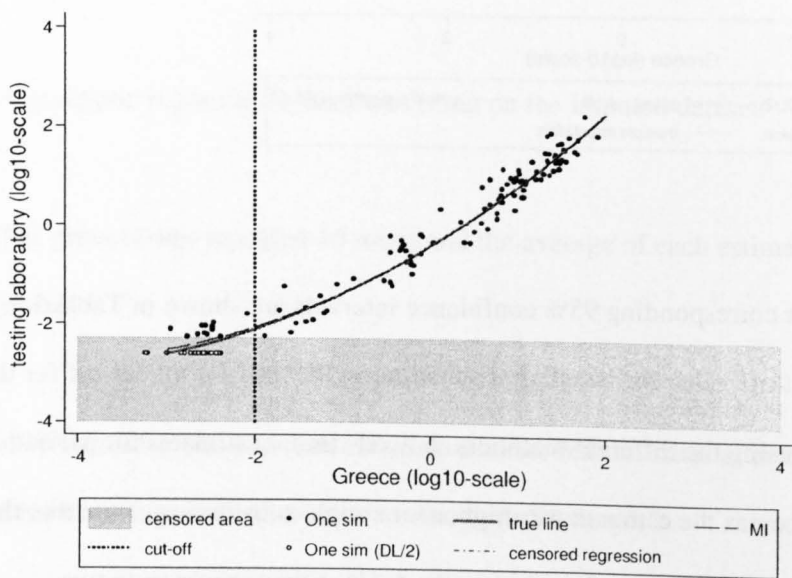
The estimates with their corresponding 95% confidence intervals are shown in Table 5.3. A comparison of the cut-off value for Israel after substituting the 0.01 IU/ml cut-off for the Greek reference centre using the different methods, showed similar estimates for MI and censored regression, whereas the estimate was higher for simple substitution. Note that the MI confidence intervals are wider than the other methods since they incorporate two different sources of variation, the variability between- and within-imputations.

Table 5.3 Standardisation equation estimates for Israel HAV panel using simple substitution, censored regression and MI (95% CIs in brackets)

method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	cut-off
substitution with $D_L / 2$	-0.39 (-0.45, -0.33)	1.10 (1.07, 1.14)	0.10 (0.08, 0.12)	-2.19
censored regression	-0.33 (-0.39, -0.26)	1.13 (1.09, 1.17)	0.04 (0.01, 0.07)	-2.43
multiple imputation	-0.32 (-0.80, 0.15)	1.12 (0.65, 1.60)	0.04 (-0.43, 0.52)	-2.40

The different methods were also compared using simulations. As above, random data were generated 1,000 times from a “true model” $y = -0.4 + 1.1x + 0.12x^2$ with a standard deviation $\sigma = 0.26$. The regression estimates were then obtained using the censored regression and MI methods. The regression equations can be viewed in Figure 5.7 below. The equations as estimated by the two methods are close to the true model.

Figure 5.7 Comparison between censored regression and MI methods from 1,000 simulations generated from $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$



Note: for graphical purposes one set of simulated results was plotted

Table 5.4 shows the mean parameter estimates following the simulations using simple substitution, censored regression and MI methods, together with the 95% coverage probabilities. The estimated cut-off value for Israel is also reported together with the 95% percentile intervals. The parameter estimates were all close to the true model. The cut-off estimate was slightly higher for the MI method, however, the 95% percentile interval included the true estimate. Note that the MI gave 100% coverage probability due to the wider confidence intervals.

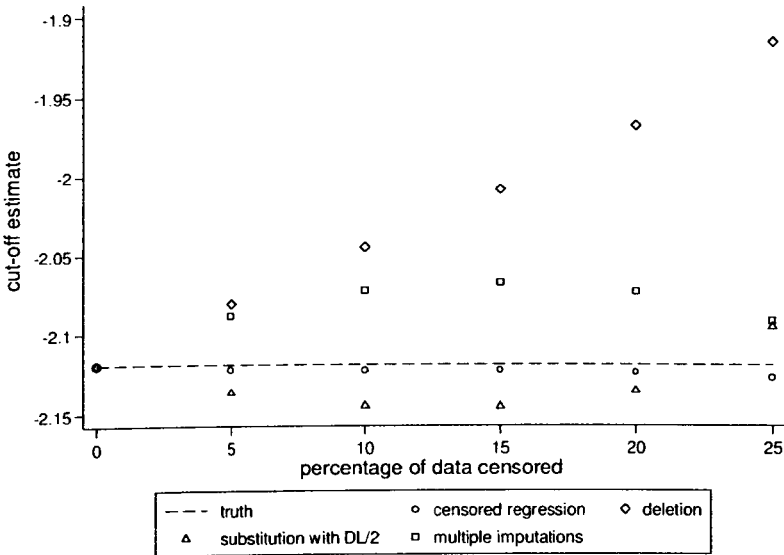
Table 5.4 Parameter estimates and 95% coverage probabilities generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$ using censored regression and MI

method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	cut-off*
true model	-0.40	1.10	0.12	-2.12
substitution with $D_L / 2$	-0.41 (93.2%)	1.11 (93.0%)	0.12 (94.8%)	-2.13 (-2.19, -2.06)
censored regression	-0.40 (94.2%)	1.10 (93.9%)	0.12 (95.1%)	-2.12 (-2.21, -2.05)
multiple imputation	-0.40 (100.0%)	1.09 (100.0%)	0.13 (100.0%)	-2.07 (-2.17, -1.99)

* 2.5th and 97.5th percentiles given instead of coverage probabilities

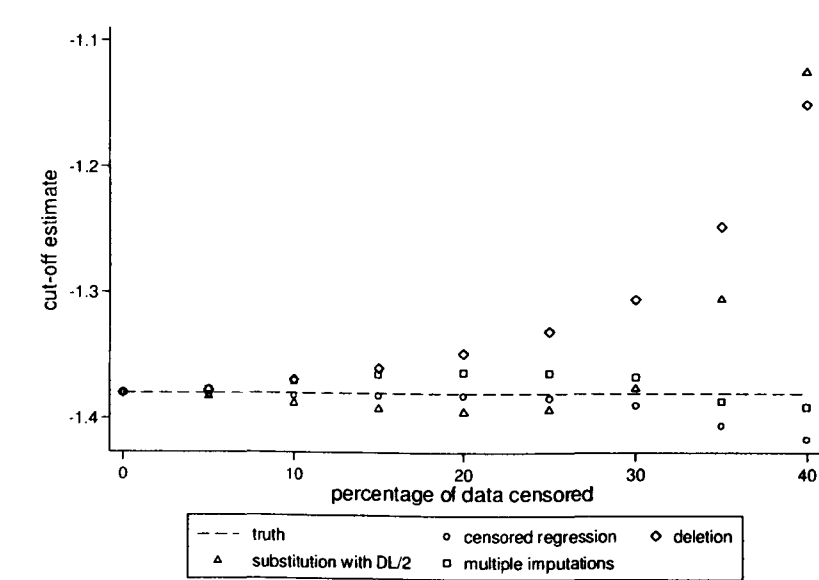
The cut-off value as estimated by the different methods is shown in Figure 5.8 for different proportions of data censored. Censored regression gave cut-off estimates closest to the truth followed by simple substitution and the MI imputation methods.

Figure 5.8 Comparison of cut-off estimate between the censored regression and MI methods for different proportions of data censored generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$



Simulations were generated after changing the reference assay cut-off from -2 to -1 (on the $\log_{10}(\text{IU/ml})$ scale) aiming to censor a higher proportion of data. Substituting this to the true model $y = -0.4 + 1.1x + 0.12x^2$ gave a cut-off of $y(-1) = -1.38$. For up to 30% of the data censored, censored regression, simple substitution (with $D_L / 2$) and MI all gave cut-off estimates close to the truth, with the first two being slightly better. For 35% to 40% of the data censored, the multiple imputation method provided the least biased estimate followed by censored regression, whereas the simple substitution and deletion methods produced estimates far away from the true cut-off (Figure 5.9). These findings agree with the literature, which suggests that simple substitution methods produce biased estimates when more than 25% of the data is censored (Lubin JH *et al.*, 2004).

Figure 5.9 Comparison of cut-off estimate between the censored regression and MI methods for different proportions of data censored generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$ using a cut-off of -1



5.10 Investigation of regression assumptions

The methods shown here rely to some extent on the assumptions of multiple linear regression.

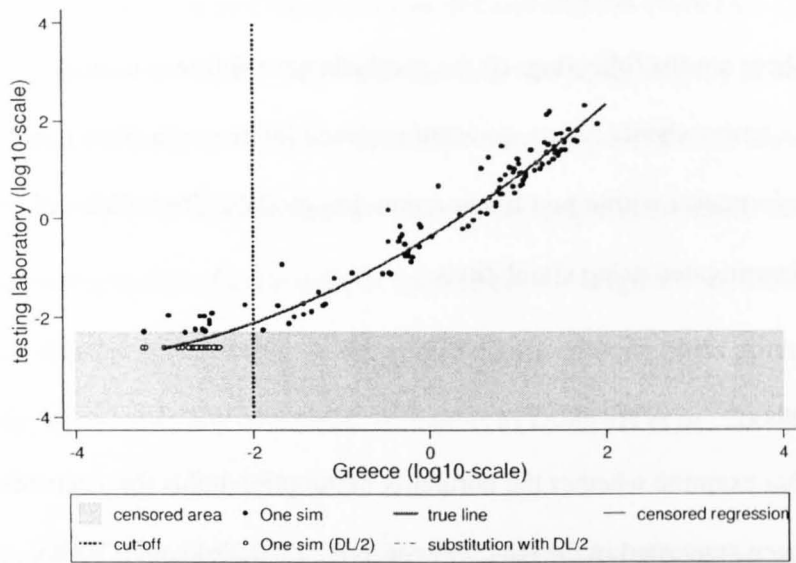
The most important of these are the following: (i) the residuals must follow a normal distribution, (ii) the error term ε should have a constant variance (homoscedasticity), and (iii) a further issue is performance within non-linear regression models. The validity of these assumptions was investigated using simulations.

(i) Normality assumption

A number of statistics that examine whether the normality assumption holds for a censored regression model have been suggested in the past (Greene WH, 2003; Holden D, 2004). However, for this Chapter, robustness to the normality assumption was investigated by simulating data using a non-normal residual distribution. Specifically, a gamma distribution $\Gamma(\lambda, \tau)$ was fitted for the distribution of residuals, where λ was defined as the shape parameter and τ as the scale parameter. Note that the distribution was shifted to centralise around zero by subtracting its mean $\lambda\tau$.

Simulations were generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = \sqrt{\lambda\tau^2} = 0.26$ and $\lambda = 2$ for the gamma residual distribution. A comparison between the simple substitution and the censored regression methods can be viewed in Figure 5.10. Both curves are close to the true line.

Figure 5.10 Comparison between simple substitution and censored regression methods from 1,000 simulations generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$ when the normality assumption does not hold ($\lambda = 2$)



Note: for graphical purposes one set of simulated results was plotted

Table 5.5 shows the estimated parameters for the simple substitution, censored regression and MI methods after 1,000 simulations. All methods produced cut-off estimates close to the true model’s cut-off with the MI method being slightly more biased.

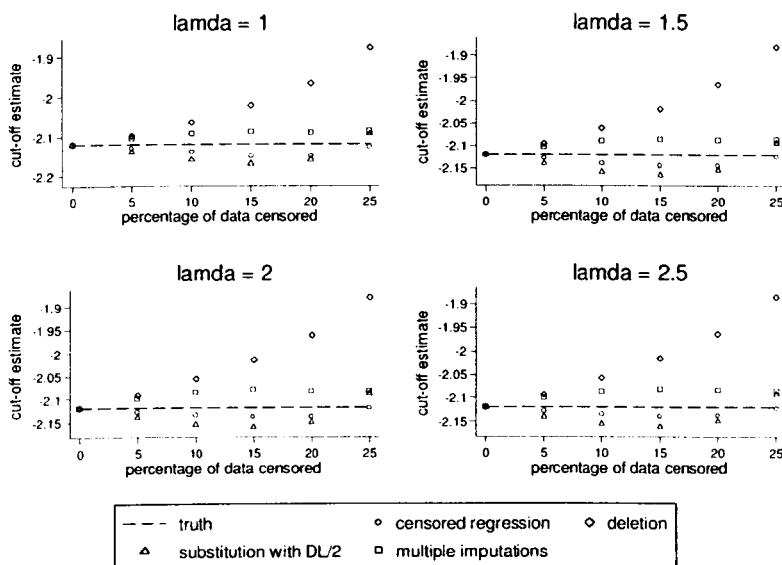
Table 5.5 Parameter estimates and 95% coverage probabilities generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$ using simple substitution and censored regression when the normality assumption does not hold ($\lambda = 2$)

method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	cut-off
true model	-0.40	1.10	0.12	-2.12
substitution with $D_L / 2$	-0.41 (92.5%)	1.11 (93.1%)	0.12 (94.7%)	-2.13 (-2.20, -2.07)
censored regression	-0.40 (94.6%)	1.10 (95.4%)	0.12 (93.3%)	-2.13 (-2.22, -2.06)
multiple imputations	-0.40 (100.0%)	1.09 (100.0%)	0.12 (100.0%)	-2.08 (-2.19, -1.99)

* 2.5th and 97.5th percentiles given instead of coverage probabilities

Figure 5.11 shows the cut-off value estimate for different proportions of data censored for that scenario after varying the shape parameter. The MI, censored regression and $D_L / 2$ methods, all gave cut-off estimates close to the truth. The robustness of the censored regression method for small deviations from the normality assumption has been demonstrated in a previous study (Lynn HS, 2001).

Figure 5.11 Comparison of cut-off estimate between censored regression and MI methods for different proportions of data censored generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.26$ when normality assumption does not hold (varying λ)



(ii) Homoscedasticity

As mentioned above, one of the assumptions of censored regression is the homogeneity of the variance (homoscedasticity). However, there may be an additional source of uncertainty that only applies to values outside the assay DL, and more often to low concentrations. This may lead to higher residual variability for these measurements (heteroscedasticity), and therefore, violation of the assumptions for censored regression.

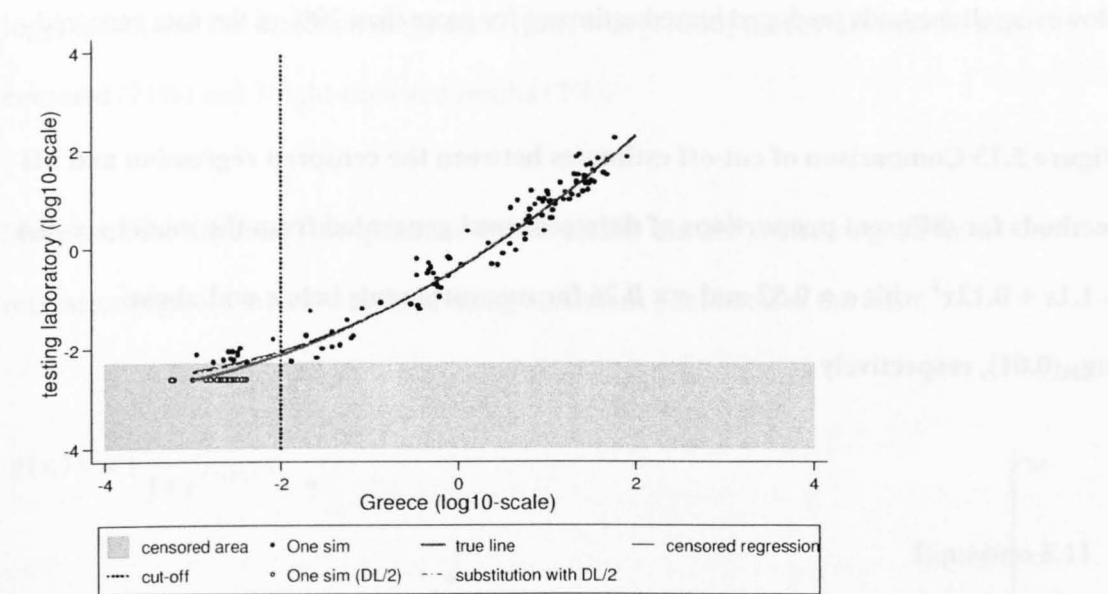
Although the impact of heteroscedasticity on censored regression estimators has been investigated in the past, analyses have focused on very specific models. The proportion of data censored seems to be one of the main factors determining the degree of estimators' bias (Greene WH, 2003).

One way to take heteroscedasticity into account is to treat measurements below and above DL differently by fitting “hybrid” models, such as the hybrid lognormal, that produce different error variances depending on whether the observation is within the DL or not. However, to construct such a model, it is essential to initially establish whether the same error variance exists across the whole distribution of the assay. This may be provided – although not always possible – by obtaining quantitative results for some measurements outside the DL (Daniels RD & Yiin JH, 2006; Whitcomb BW & Schisterman EF, 2008).

The effect of heteroscedasticity on the regression estimates was examined as follows: The different methods of handling censored data were evaluated after doubling the amount of variation for lower concentrations. Using the same example as above, data were generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with a standard deviation $\sigma = 0.26$ for measurements above the reference cut-off value of $\log_{10}(0.01)$ and $\sigma = 0.52$ for measurements below $\log_{10}(0.01)$.

Figure 5.12 shows the estimated curves using the simple substitution and the censored regression models after 1,000 simulations. Although both curves are very close to the true model there is some discrepancy at the lower end of the equation where there is higher amount of variability.

Figure 5.12 Comparison between simple substitution and censored regression methods from 1,000 simulations generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.52$ and $\sigma = 0.26$ for measurements below and above $\log_{10}(0.01)$, respectively



Note: for graphical purposes one set of simulated results was plotted

Table 5.6 shows a comparison between simple substitution, censored regression and MI estimates following simulations. The simple substitution and censored regression methods produced cut-off estimates closer to the truth.

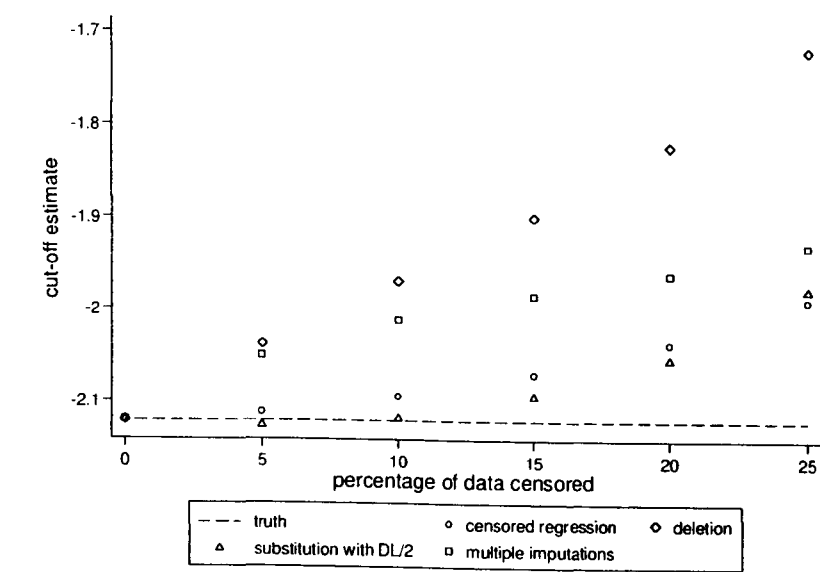
Table 5.6 Parameter estimates and 95% coverage probabilities generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.52$ and $\sigma = 0.26$ for measurements below and above $\log_{10}(0.01)$, respectively

method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	cut-off*
true model	-0.40	1.10	0.12	-2.12
substitution with $DL_L / 2$	-0.42 (95.3%)	1.09 (95.5%)	0.13 (83.8%)	-2.06 (-2.14, -1.98)
censored regression	-0.41 (95.3%)	1.09 (93.4%)	0.14 (79.3%)	-2.04 (-2.13, -1.96)
multiple imputations	-0.41 (100%)	1.07 (100%)	0.15 (100%)	-1.97 (-2.07, -1.88)

* 2.5th and 97.5th percentiles given instead of coverage probabilities

The results of the simulations, presented in Figure 5.13 below, show that simple substitution and censored regression produced the closest cut-off estimates to the truth. However, all methods produced biased estimates for more than 20% of the data censored.

Figure 5.13 Comparison of cut-off estimates between the censored regression and MI methods for different proportions of data censored generated from the model $y = -0.4 + 1.1x + 0.12x^2$ with $\sigma = 0.52$ and $\sigma = 0.26$ for measurements below and above $\log_{10}(0.01)$, respectively



(iii) Non-linear regression models

Different types of equations were used to describe the relationship between the reference centre’s and the testing laboratory’s standardisation panel results. In some cases, however, straight line or quadratic models failed to provide a good fit around the critical area of the positive / negative cut-off.

The example below demonstrates the use of a non-linear regression model. For this example the Cypriot rubella panel results were plotted against the German reference centre results. The positive / negative cut-off for the reference centre was 4 IU/ml (0.6 on the \log_{10} -scale), and the detection range for Cyprus was [1, 200] (in IU/ml) with 31 left-censored (21%) and 3 right-censored results (2%).

Since both the linear and the quadratic curves clearly did not provide a good fit, a regression sigmoid model was fitted using maximum likelihood of the form:

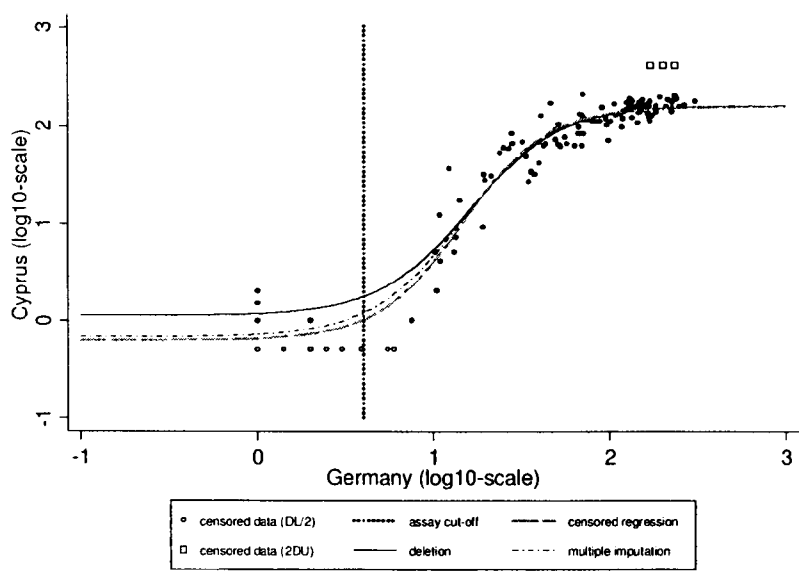
$$g(x_i) = \alpha + \frac{\beta}{1 + e^{-(\gamma + \delta x_i)}},$$

Equation 5.11

where $i = 1, \dots, 149$ denotes the panel samples and α, β, γ and δ are the model parameters. Using the ESEN2 standardisation methodology, regression based on deletion was chosen to obtain the standardisation equation, given the large differences between the two lines on the cut-off point ($COR = 0.099$).

After substituting Equation 5.11, it is possible to obtain estimates for the equivalent regression type after taking censoring into account. Model estimates using the different methods are shown in Figure 5.14. For higher measurements the different models gave similar results. However, for less than 1 \log_{10} (reference IU/ml), the deletion method gave higher results than censored regression and MI methods since it ignored the potential effect of left-censored data.

Figure 5.14 Non-linear regression curves using deletion, censored regression and MI



The parameter estimates with their corresponding 95% confidence intervals are given in Table 5.7. The method of deletion produced very different regression estimates from the other methods. The cut-off estimate ($\hat{y}_c(0.6) = 0.24$) was much higher than for the other methods.

Table 5.7 Standardisation equation estimates using simple substitution and deletion methods (95% CIs in brackets)

method	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$	cut-off
	0.05	2.14	-4.68	3.90	
deletion	(-0.07, 0.18)	(1.99, 2.29)	(-5.64, -3.72)	(3.15, 4.65)	0.24
substitution with $D_L / 2$	-0.27	2.45	-5.05	4.36	
and $2D_U$	(-0.34, -0.20)	(2.35, 2.55)	(-5.09, -4.19)	(3.63, 5.09)	-0.07
	-0.20	2.38	-4.86	4.19	
censored regression	(-0.30, -0.11)	(2.26, 2.50)	(-5.71, -4.01)	(3.49, 4.89)	0.01
	-0.17	2.35	-4.50	3.90	
multiple imputation	(-0.51, 0.18)	(2.00, 2.70)	(-5.35, -3.68)	(3.27, 4.54)	0.07

The methods used were validated through simulations. As before, random data were generated 1,000 times using the reference panel results. The true model was set as

$$y = -0.3 + \frac{2.5}{1 + \exp(-(-4) - 3.5x)}$$

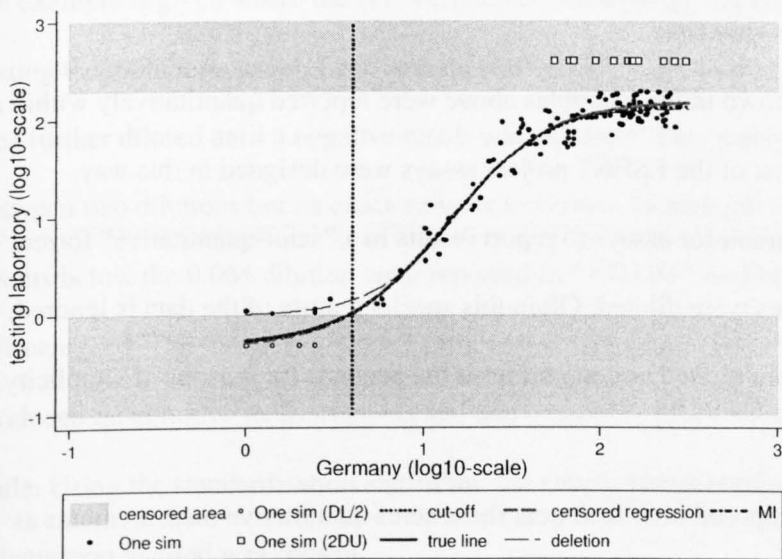
with a standard deviation $\sigma = 0.16$. Substituting a cut-off of

0.6 to the true model gave $y(0.6) = 0.03$. Following each simulation, regression equations were estimated using the methods of deletion, simple substitution, censored regression and MI after censoring all measurements below $\log_{10}(1)$ and above $\log_{10}(200)$ IU/ml.

Figure 5.15 shows how the models based on deletion, censored regression and MI compare to the true model the data was generated from. The censored regression and MI curves remain very close to the true line, whereas the model based on deletion deviates at the end of the lines, especially at the lower end where there is a higher number of left-censored data.

Figure 5.15 Comparison between censored regression and the deletion methods after

1,000 simulations generated from $y = -0.3 + \frac{2.5}{1 + \exp(-(-4) - 3.5x)}$ with $\sigma = 0.16$



Note: for graphical purposes one set of simulated results was plotted

The parameter estimates for the different methods are shown in Table 5.8. The cut-off estimate was close to the truth for all methods with exception of the deletion method ($\hat{y}_c^*(0.6) = 0.16$).

Table 5.8 Parameter estimates and 95 % coverage probabilities generated from the model $y = -0.3 + \frac{2.5}{1 + \exp(-(-4) - 3.5x)}$ with $\sigma = 0.16$ using simple substitution, censored regression and MI methods

method	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\delta}$	cut-off*
true model	-0.30	2.50	-4.00	3.50	0.03
deletion	-0.00 (10.1%)	2.13 (7.9%)	-5.07 (42.9%)	4.21 (51.5%)	0.16 (0.06, 0.24)
substitution with	-0.35	2.59	-3.93	3.42	0.00
$D_L / 2$ and $2D_U$	(93.5%)	(81.2%)	(95.5%)	(94.4%)	(-0.07, 0.08)
censored	-0.33	2.53	-3.96	3.48	0.02
regression	(96.4%)	(95.9%)	(95.6%)	(94.9%)	(-0.08, 0.11)
multiple imputation	-0.30 (97.7%)	2.47 (98.0%)	-4.03 (67.0%)	3.57 (75.2%)	0.03 (-0.07, 0.12)

* 2.5th and 97.5th percentiles given instead of coverage probabilities

5.11 Interval censored regression

The serological results shown in the examples above were reported quantitatively within a detection range, since most of the ESEN2 project assays were designed in this way. However, it is not uncommon for assays to report results in a “semi-quantitative” format due to the way the samples were diluted. Often this special feature of the data is ignored, and such measurements are treated as continuous in the analysis for reasons of simplicity.

An alternative method proposed here is to treat these semi-quantitative measurements as interval censored data. Then appropriate techniques, such as the *interval censored*

regression method, can be applied which allow for bias adjustment generated from this type of data (Lyles RH *et al.*, 2001; Zhang Z & Sun J, 2010).

Suppose that a sample of size $N + n$ consists of a set of m intervals with sizes N_1, N_2, \dots, N_m $\left(\sum_{i=1}^m N_i = N \right)$. Then a measurement y_{ij} , where $i = 1, \dots, m$ and $j = 1, \dots, N_i$, belongs to the interval $[\min(y_i), \max(y_i)]$. Also, let $N + 1, \dots, c$ be left-censored data and $c + 1, \dots, n$ be right-censored data.

Then the log-likelihood shown in Equation 5.4 can be generalised as:

$$\begin{aligned} \ln(L) = & \sum_{i=1}^m \sum_{j=1}^{N_i} \ln \left[F(\max(y_i); g(x_{ij}), \sigma) - F(\min(y_i); g(x_{ij}), \sigma) \right] + \\ & + \sum_{k=N+1}^c \ln \left[F(y_k; g(x_k), \sigma) \right] + \sum_{k=c+1}^n \ln \left[1 - F(y_k; g(x_k), \sigma) \right]. \end{aligned}$$

Equation 5.12

An example is given where the NT-Vero assay was used by the Finnish laboratory for testing the diphtheria panel. Each sample was tested by dilution and, if found positive, it was further diluted until a negative result was obtained. This means that the titre lies between two dilutions but its exact value is unknown. In addition to the interval censoring, results below the 0.004 dilution were reported as “< 0.004” and hence, were treated as left-censored. The Italian reference centre tested their panel with DA-DELFI which produced quantitative measurements and had a cut-off of 0.01 IU/ml or -2 on the \log_{10} -scale. Using the standardisation algorithm, the simple linear regression based on the simple substitution method was chosen.

Figure 5.16 and Table 5.9 show the comparison between the regression equations when ignoring interval censoring i.e. using deletion and simple substitution, and when adjusting for it using an interval censored regression model. The interval censored regression method gave a slightly higher slope, and therefore, a higher cut-off estimate ($\hat{y}(-2) = -2.36$).

Figure 5.16 Comparison between deletion, simple substitution and interval censored regression methods

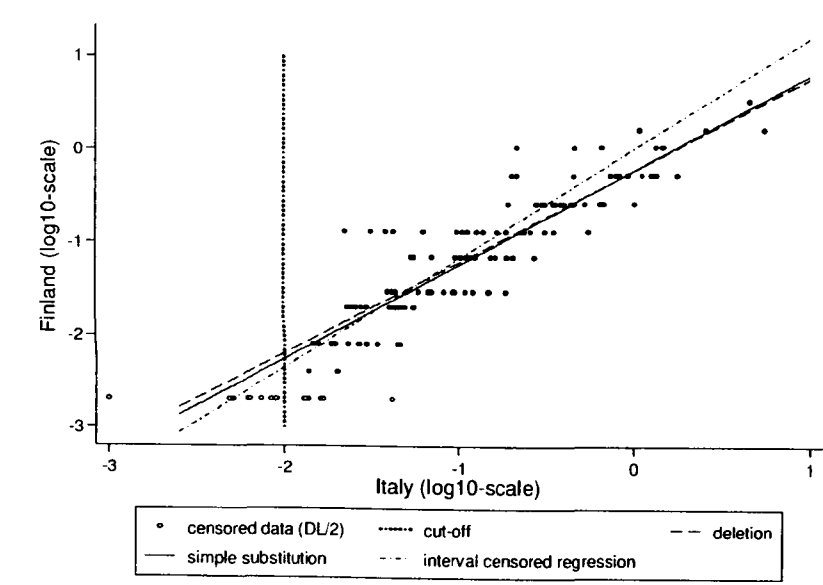


Table 5.9 Model estimates using deletion, simple substitution and interval censored regression methods (95% CIs in brackets)

method	$\hat{\beta}_0$	$\hat{\beta}_1$	cut-off
deletion	-0.24 (-0.32, -0.15)	0.98 (0.89, 1.07)	-2.20
substitution with $D_L / 2$	-0.23 (-0.32, -0.15)	1.01 (0.95, 1.08)	-2.25
interval censored regression	-0.00 (-0.10, 0.09)	1.18 (1.09, 1.26)	-2.36

The different methods were compared using simulations. Once measurements for the Finnish laboratory were generated from the true model, they were divided into groups to resemble serological data from dilution series (Table 5.10). Any measurements ≤ 0.004 or

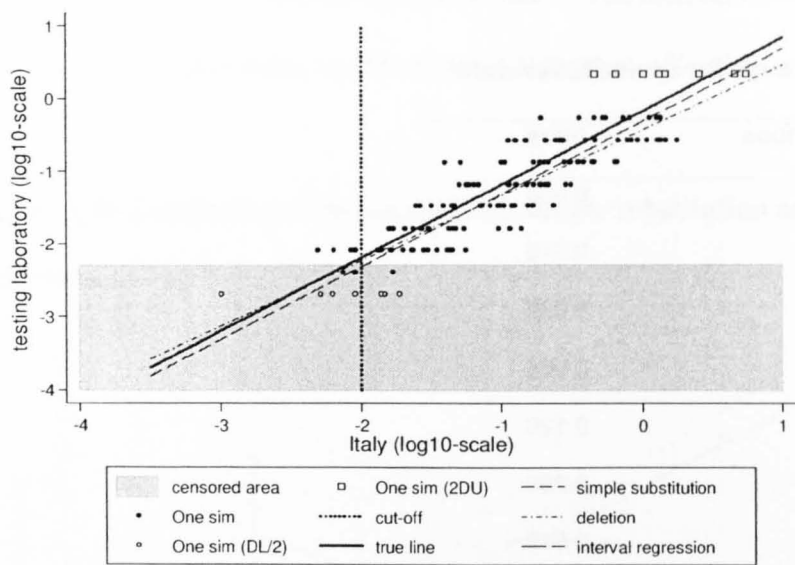
> 1.024 were treated as censored. Note that the data were \log_{10} -transformed prior to the analysis.

Table 5.10 Dilution groups for quantitative data

quantitative data groups	titres
(0.004, 0.008]	0.008
(0.008, 0.016]	0.016
(0.016, 0.032]	0.032
(0.032, 0.064]	0.064
(0.064, 0.128]	0.128
(0.128, 0.256]	0.256
(0.256, 0.512]	0.512
(0.512, 1.024]	1.024

Figure 5.17 shows the random data that were generated from a true model $y = -0.2 + x$ with $\sigma = 0.3$ and classified into dilutions after 1,000 simulations (the results of one simulation are shown). The regression estimates were obtained using deletion, simple substitution and interval censored regression methods. The interval censored regression seems closer to the true line than the other two methods.

Figure 5.17 Comparison between deletion, simple substitution and interval censored regression methods after 1,000 simulations from $y = -0.2 + x$ with $\sigma = 0.3$



Note: for graphical purposes one set of simulated results was plotted

Table 5.11 shows the model estimates using deletion, simple substitution and interval censored regression together with their 95% coverage probabilities. The estimated cut-off value for Finland was also reported together with the 95% percentile intervals. The interval censored regression method produced a cut-off estimate closer to the truth than the other methods.

Table 5.11 Parameter estimates and 95% coverage probabilities generated from the model $y = -0.2 + x$ with $\sigma = 0.3$

method	$\hat{\beta}_0$	$\hat{\beta}_1$	cut-off*
true model	-0.20	1.00	-2.20
deletion	-0.46 (0.1%)	0.89 (23.5%)	-2.24 (-2.33, -2.16)
simple substitution	-0.34 (13.6%)	0.99 (95.2%)	-2.32 (-2.40, -2.24)
interval censored regression	-0.20 (95.9%)	1.00 (95.2%)	-2.20 (-2.30, -2.10)

* 2.5th and 97.5th percentiles given instead of coverage probabilities

5.12 Method comparison when censoring occurs on the independent variable

During the ESEN2 project, laboratories were asked to provide quantitative results when testing the standardisation panel. However, this was not always possible and some laboratories reported censored data. The reference centres provided quantitative results for all antigens apart from VZV where a small number of samples were reported as below the assay DL. For such a scenario the methods of deletion and simple substitution were used as described above, despite the fact that the censoring occurred on the independent variable (Kafatos G *et al.*, 2005).

Non-parametric methods have been proposed for dealing with x-censored data based on Kendall's tau statistic. Kendall's line is obtained by calculating the median of all possible pairwise slopes. Although this method has since been further developed and today a number of variations exist, the disadvantage, apart from being computer intensive, is that it only applies for straight lines (Akritas MG *et al.*, 1995; Helsel DR, 2005; Lim J, 2006).

As an alternative to Kendall's line, a non-parametric test is proposed here to decide whether the regression equation estimates after excluding the x-censored data can be considered unbiased. The method can be described as follows:

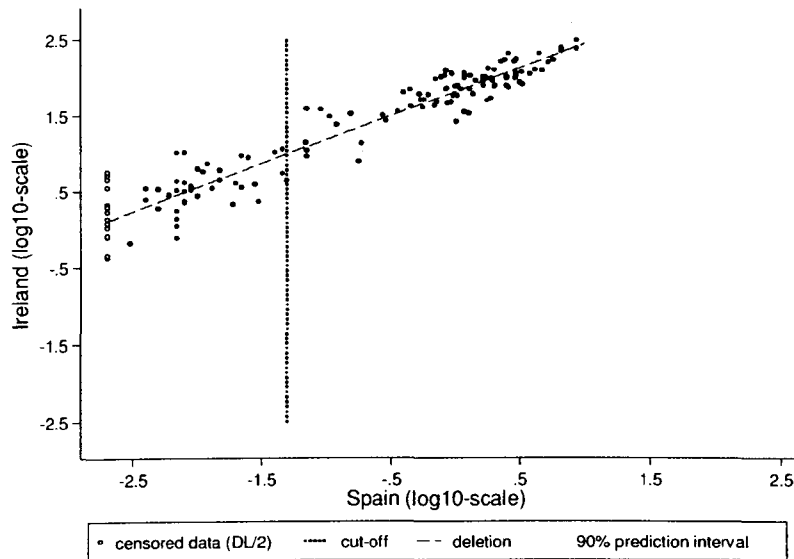
- (i) Fit a regression equation of the type shown in Equation 4.1 based only on the quantitative data (i.e. omit the x-censored values).
- (ii) Obtain the 90% *prediction interval (PI)* for y at the point where x-values are censored.
- (iii) Test whether the number of censored x-values outside the 90% PI is consistent with the expected number following a binomial distribution.

If the number of censored x-values is within what would be expected, then the regression model based on the deletion method can be considered valid.

An example is given where the Irish VZV standardisation panel of 148 samples is plotted against the Spanish reference centre's results. The Irish lab reported all-quantitative results, whereas the Spanish reference centre reported 18 samples (12%) below 0.002 IU/ml. The positive / negative cut-off was 0.05 IU/ml (-1.3 on the \log_{10} -scale) as given by the reference centre.

Figure 5.18 shows the fitted line $\hat{y}_i = 1.81 + 0.63x_i$ ($i = 1, \dots, 148$) together with the 90% PI. There are 5 samples above the 90% PI. Given that there are 18 censored observations for x, it would be expected that $18 \times 0.05 = 0.9$ samples would be above the 90% prediction interval. Using a binomial test, the observed censored data above the prediction limits (5/18) are significantly different than those expected ($p = 0.002$), which suggests a potential change of direction for the fitted line below DL. Therefore, the fitted line using the deletion method may not be appropriate to use as the standardisation equation in this lower range.

Figure 5.18 Regression curves using deletion and 90% PIs for a panel comparison with x-censoring



The idea behind this nonparametric method, is to test whether fitting that particular type of equation (linear, quadratic or other) based on the deletion method, is reasonable for the censored dataset shown. Given that the x-variable is censored, it is not possible to know the shape of the regression equation in the censored area, i.e. whether it follows the same equation as in the area where quantitative measurements are available.

By testing whether the number of samples above the prediction interval is within normal expectations, the possibility that the data in the censored area follow a different type of equation is examined. Note that for the above example, the main interest is in the data above the prediction interval (and not so much below), because these are more likely to be inconsistent with the rest of the equation. There are two main possibilities for the x-censored values: either there is a linear relationship or the data level-off. The latter would mean that a different type of equation could be more suitable than the linear, since it would fit better the y-values in the censored area. There is a third possibility of generally higher

variability at low levels. In this case we would expect similar numbers of x-censored data below and above the prediction intervals.

5.13 Conclusions

The results of the simulations showed that for up to 20 - 25% of censored data, simple substitution and deletion methods generally gave estimates close to the truth at the crucial point of positive / negative cut-off. This is probably sufficient for the panel test comparisons carried out during the ESEN2 project, since small proportions of data were reported as censored in most cases. However, in scenarios with more than 25% censored observations, the simple substitution and deletion estimates become biased whereas censored regression and MI methods continue to give accurate estimates (Lubin JH *et al.*, 2004).

Although the simulations confirm that the methods used in the ESEN2 project to take into account censoring were valid, a case has been made for using the censored regression method instead in the future. The censored regression method can be easily applied, since it is readily available in most statistical software (cnreg command in Stata), although it may be necessary to write a maximum likelihood routine for non-linear regression models (examples of such routines are given in Appendix I(A)). When regression assumptions were violated, censored regression continued to give acceptable cut-off estimates as shown by the simulation examples.

The MI method has the advantage of being a more robust method for high proportions of censored data (>25%). However, it is often not a straight-forward method to implement, and some variations of this method can be computer intensive. The usage of censored regression is also advised by Lubin (amongst others), who claims that "...multiple

imputation is necessary only if explicit values are needed for measurements below DL” (Lubin JH *et al.*, 2004).

Two novel methods were proposed for different types of censored data. Interval censored regression models were suggested for assays with dilution series. In the simulation example shown, the interval censored regression method gave a cut-off estimate closest to the truth. For x-censored data, a non-parametric method was proposed for testing whether the regression model estimates using the deletion method can be considered valid to describe the relationship between the assay tests.

Another method which was not examined in this Chapter includes the replacement of measurements below the DL with a constant equal to the expectation of the censored data, in order to obtain unbiased estimates (Akritas MG *et al.*, 1995; Richardson DB & Ciampi A, 2003; Schisterman EF *et al.*, 2006; Lubin JH *et al.*, 2004). Substitution with $E(Y_i | Y_i < D_L)$ is easy to implement, and it has been found in the past to produce better estimates than the simple substitution with $D_L / 2$ (Lynn HS, 2001). However, the drawback of substituting with $E(Y_i | Y_i < D_L)$ is its dependency on distributional assumptions (Lubin JH *et al.*, 2004).

As mentioned above, there have been a number of papers during the last 5 years examining different methods to account for censored observations when estimating regression coefficients. Although some of the methods described in this Chapter are not new, it was important to apply them to these particular sets of serological data, since they have been found in the past to give different results under different scenarios.

In conclusion, the simple substitution and deletion methods as presented in the previous Chapter, and used for the ESEN2 project, seem to be working satisfactorily. However, for a future ESEN project, censored regression could provide better estimates.

Chapter 6: Comparison between standardised and non-standardised serological results

6.1 Serosurvey construction and national sero-profiles

As part of the ESEN2 project, each participant laboratory was asked to collect and test its own national serosurvey. Each country was advised to collect around 3,500 samples stratified by age and sex. In this way, at least 200 samples would be obtained for each age group, a number which, according to the sample size calculations, would mean a 95% confidence interval width of 4% to 6% for seroprevalence estimates.

In practice, smaller serosurveys were collected of minimum size 1,000 - 1,500 samples or 100 samples per year stratum (gender was occasionally ignored). Although a population-based serosurvey would be ideal to minimise selection bias, this was only carried out by a few countries due to logistical difficulties and high expense. Most participant countries collected serum specimens from residues remaining after the completion of microbiological or biochemical investigations (Osborne K *et al.*, 2000; Osborne K *et al.*, 1997). However, the representativeness of national serosurveys is beyond the scope of this thesis.

Following the collection of the serosurvey, each national laboratory tested the samples using an assay of their choice. During the ESEN2 project, a standardisation algorithm was developed as a means to compare serological results tested in different national laboratories (Chapter 4). This was necessary in order to overcome the issue of variability between assays and laboratory methods (Andrews N *et al.*, 2000; Kafatos G *et al.*, 2005).

Once the serosurvey results had been standardised, seroprevalence was estimated for each country by age group (Andrews N *et al.*, 2000; Kafatos G *et al.*, 2005). Theoretically, any discrepancies between standardised and non-standardised results could be explained within the context of variability, due to the variety of laboratory techniques and assays (Chapters 2 and 4).

The main aim of this Chapter is to assess the impact of standardisation on population seroprevalence or national *sero-profiles*. The hypothesis is that if the bias corrected by standardisation does not greatly affect seroprevalence, then there might be a case against the benefits and necessity of standardising. A second aim is to identify the source of any differences between standardised and non-standardised seroprevalences.

6.2 Classification into negative, positive and equivocal results

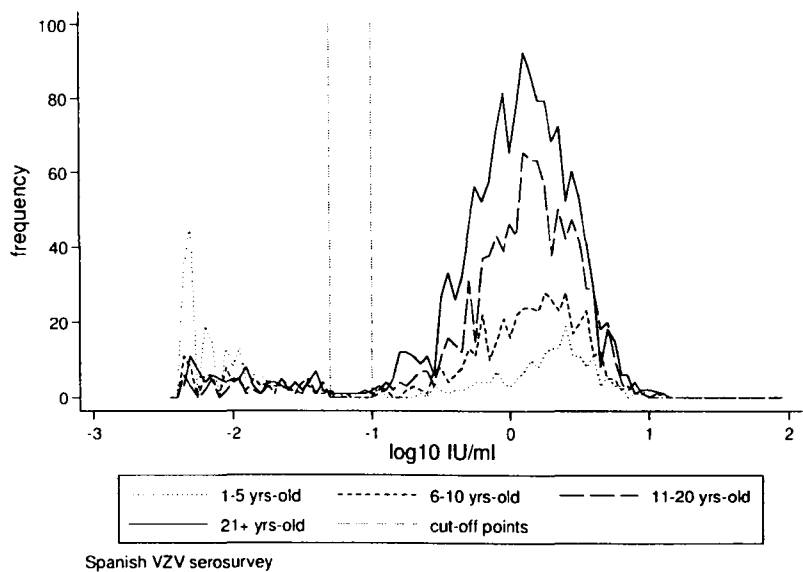
Serological assays are used primarily for individual patient diagnosis. For this purpose, a simple binary outcome, either negative (unprotected against infection) or positive (protected against infection), should be sufficient. As discussed in Chapter 2, many of these assays return a continuous outcome which can subsequently be classified into positive or negative, by applying a cut-off point in accordance with the assay manufacturer's specifications.

During the previous Chapters, it has been assumed that for each assay a unique cut-off point exists that classifies samples into positive and negative. However, this is not strictly true since many assays have two cut-off points instead of one. Using these, quantitative measurements can be classified into negative, positive and *equivocal* or *low positive* results.

In the following example, the serosurvey that was carried out by Spain in 1996 as part of the ESEN2 project is presented. The serosurvey involved collecting and testing 3,605 samples for VZV using an ELISA assay (Enzygnost VZV, Dade Behring, Germany) (de Ory F *et al.*, 2006). Two cut-off points were specified by the assay manufacturer: 0.05 and 0.1 IU/ml. According to these, 522 samples were classified as “negative”, 14 samples were classified as “equivocal” and 3,069 as “positive”. The \log_{10} -transformed distribution of

IU/ml by four different age groups together with the cut-off range can be viewed in Figure 6.1.

Figure 6.1 Frequency distribution of VZV Spanish results by age group



Since the aim of the ESEN2 project was to estimate seroprevalence in the population, a single cut-off point was required. 0.05 IU/ml was chosen i.e. the lower cut-off point specified by the assay manufacturer (de Ory F *et al.*, 2006). Merging equivocal with positive results is a practice commonly applied to serological data, and it was used for a number of different antigens and assays throughout the ESEN and ESEN2 projects (Andrews N *et al.*, 2008;de Melker H *et al.*, 2001;Pebody RG *et al.*, 2000;Tischer A *et al.*, 2007;Andrews N *et al.*, 2000). In this thesis, any future reference to a single assay cut-off will mean that such merging has taken place where necessary.

As far as the above example is concerned, there was little difference in seroprevalence estimates when the equivocal results were merged with positive or negative samples or

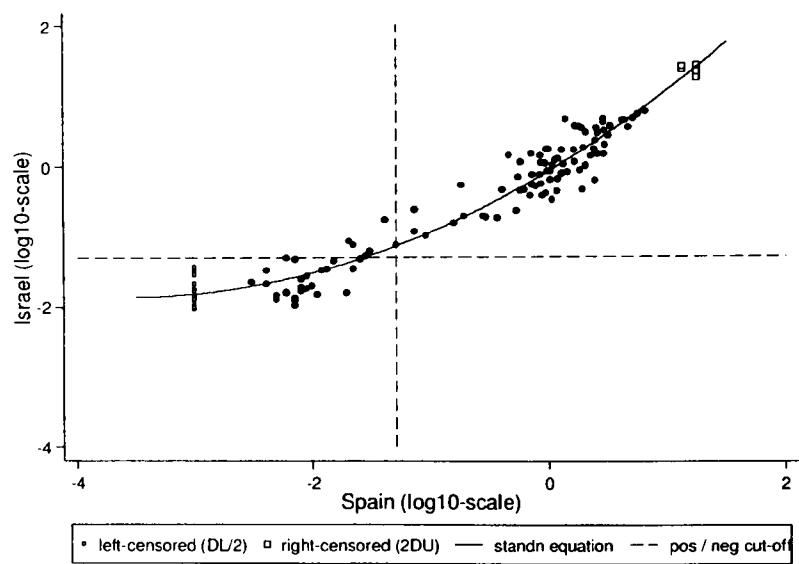
were ignored altogether, since few observations in this serosurvey were classified as equivocal.

6.3 Obtaining standardised and non-standardised sero-profiles

For the purposes of the ESEN2 project, the standardisation algorithm was applied to the serosurvey results for each country resulting in common units. These measurements were classified as either “negative” or “positive” according to the reference centre’s assay manufacturer. The national seroprevalence was then calculated by estimating the percentage of positive samples for each country and age group.

A serosurvey case of 1,592 sera was tested by the Israel national laboratory for VZV using the same assay as the Spanish reference laboratory (Enzygnost VZV, Dade Behring, Germany). After a panel of 133 samples were tested by both laboratories, a quadratic equation was selected to transform Israel’s results into comparable units according to the Spanish reference centre’s assay specifications (de Ory F *et al.*, 2006). The standardisation curve and the reference centre’s positive / negative cut-off (i.e. 0.05 IU/ml or -1.3 on the \log_{10} -scale) are given in Figure 6.2 below.

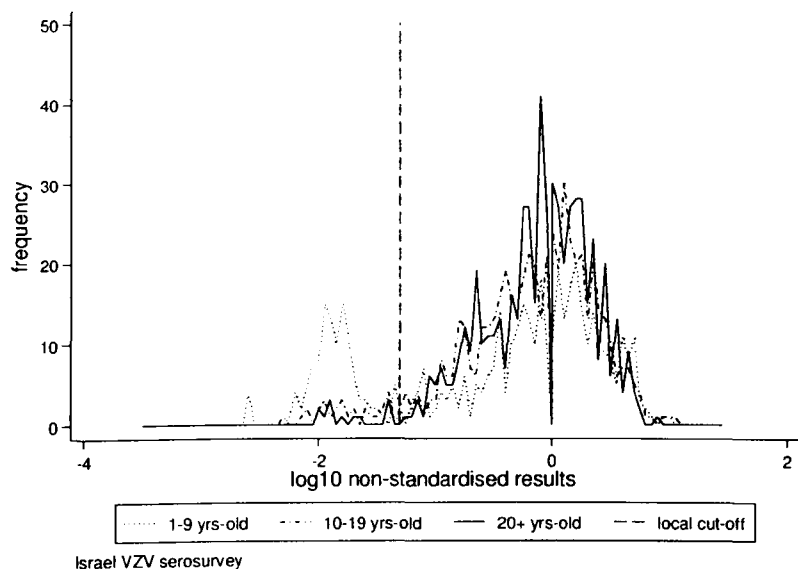
Figure 6.2 Standardisation curve for Israel VZV results



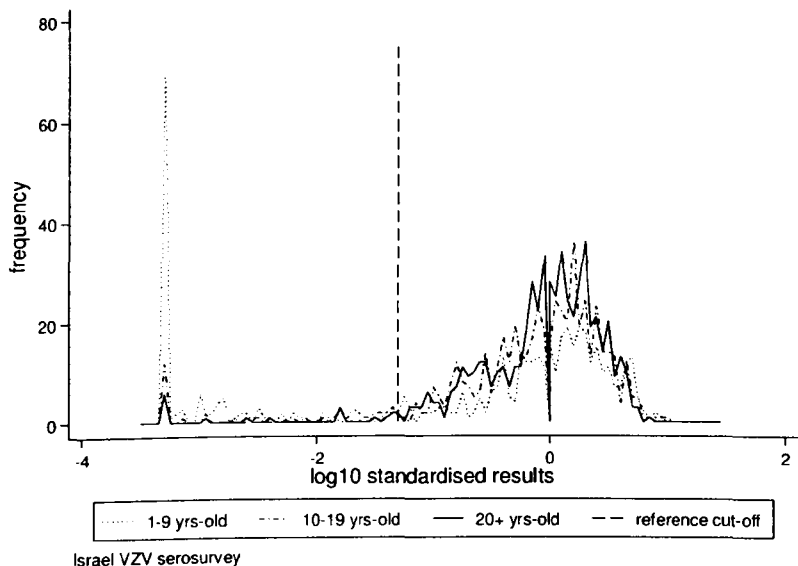
The equation $\hat{y}_i = -0.08 + x_i + 0.14x_i^2$ ($i = 1, \dots, 133$) was selected to standardise the Israel serosurvey results as shown in Equation 4.11. Figures 6.3(a) and 6.3(b) show the distribution of serological results prior to and after standardising for three age groups (1-9, 10-19 and more than 20 years).

Figure 6.3 Distribution of Israel’s VZV serosurvey results by age groups

(a) Non-standardised measurements

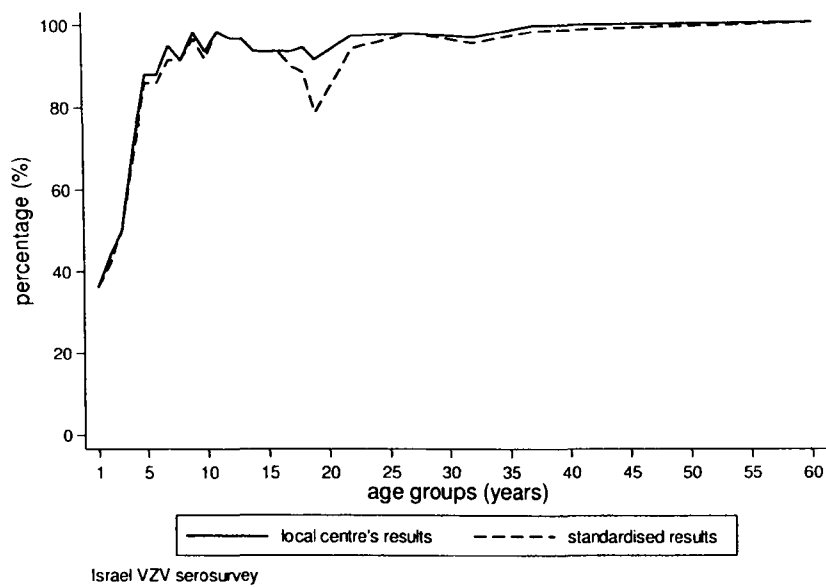


(b) Standardised measurements



Following standardisation, continuous measurements may be classified into negative or positive results. Figure 6.4 below presents a comparison between the standardised and non-standardised outcome.

Figure 6.4 Sero-profile for Israel’s VZV results before and after standardisation



The difference between standardised and non-standardised results suggests that a number of samples valued near the negative / positive cut-off point may be classified differently by the national and the reference laboratories, especially among 15-25 year olds.

A qualitative comparison of the data is presented in Table 6.1. There was a discrepancy of 23 samples (i.e. 1.5% of the data) that were tested as positive by the Israel assay but were re-classified as negatives following standardisation.

Table 6.1 Comparison between standardised and non-standardised results classified as negative and positive

		standardised classification		
		negative	positive	Total
local classification	negative	162	0	162
	positive	23	1407	1430
	Total	185	1407	1592

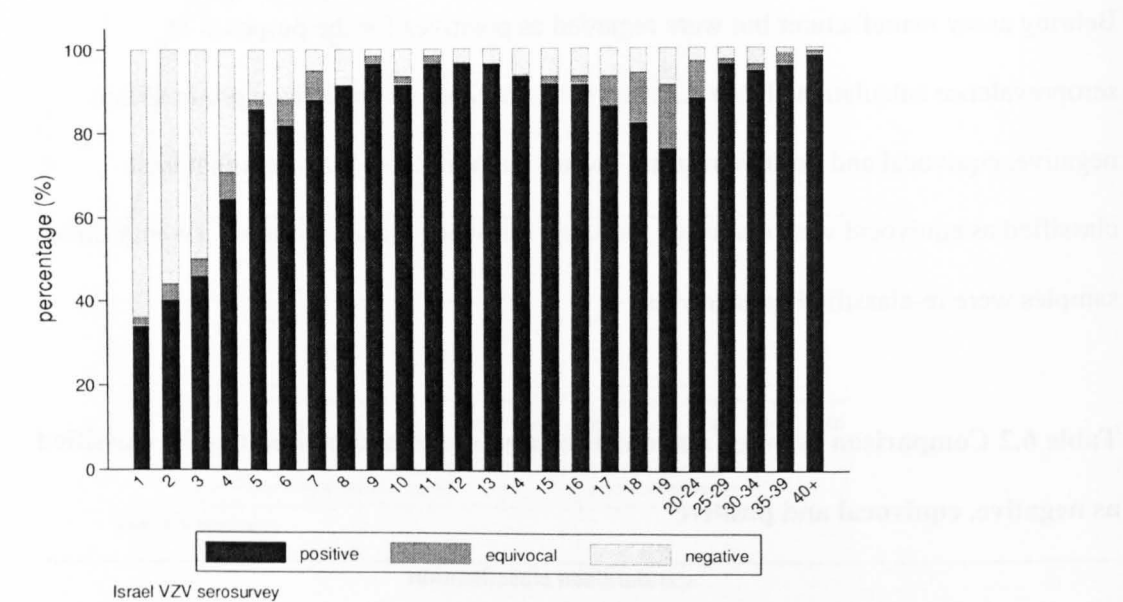
The discrepancy may be further investigated by re-introducing the “equivocal” classification. Equivocal samples were defined as 0.05 to 0.1 IU/ml according to the Dade Behring assay manufacturer but were regarded as positives for the purposes of seroprevalence calculation. Table 6.2 shows the qualitative comparison grouped into negative, equivocal and positive results. Twenty-eight of the 51 samples that were classified as equivocal were classified the same following standardisation. Twenty-three samples were re-classified as negatives.

Table 6.2 Comparison between standardised and non-standardised results classified as negative, equivocal and positive

		standardised classification			
		negative	equivocal	positive	Total
local classification	negative	162	0	0	162
	equivocal	23	28	0	51
	positive	0	21	1358	1379
	Total	185	49	1358	1592

Figure 6.5 below shows Israel’s VZV results by age, classified into negative, equivocal and positive before they were standardised. It was mainly the equivocal results for ages 17 to 24 years, of which there was a high number, that were re-classified following standardisation. Given that for Israel there was no routine vaccination programme established at the time the serosurvey was tested, it is possible that the equivocal results represent waning antibody immunity following childhood infection (Pinot de Moira A & Nardone A, 2005). Note that in this case, the standardisation helped to identify a potentially important epidemiological finding which may otherwise have been missed.

Figure 6.5 Non-standardised results for Israel VZV serosurvey by age group
classified into negative, equivocal and positive



6.4 Between-laboratory and within-laboratory variability

As was briefly mentioned in Chapter 4, variability between serological results can be categorised into the following two types:

(a) For sera that are tested within a laboratory using a specific assay, variability in test results will be referred to as *within-laboratory* variability. This may occur, either due to slight changes in the conditions in which the test took place or due to different individuals carrying out the test. Differences between assay plates may also be responsible for this variability but this will be examined in Chapter 9.

(b) For the purposes of the ESEN2 project, national serosurveys were tested in different laboratories. This introduced a different source of variation that will be referred to as *between-laboratory* variability. This type of variation may occur as laboratories use slightly different procedures and / or different assays. Since between-laboratory variability

is systematic (e.g. an assay being more sensitive than others), it can also be referred to as *bias*.

In most situations between-laboratory variability would be expected to be higher than within-laboratory variability. Standardisation provides a method of correcting such bias.

6.5 Impact of standardisation on sero-profiles

Assuming no failure of the standardisation methodology, the difference between standardised and non-standardised seroprevalence estimates serves as a measurement of the within- and mainly between-laboratory variability. In other words, the larger the bias between laboratory results the greater the difference between standardised and non-standardised outcomes.

However, the exact nature of this variability is not clear. It is very likely a combination of variability generated from different assay methods, and differences in laboratory techniques and procedures. The level of the variability may also be expected to vary between different antigens and age groups.

In order to provide a measure of the variation between standardised and non-standardised seroprevalence estimates, the difference in seroprevalence estimates across age groups, EP_j , for age groups $j = 1, \dots, \zeta$, was defined as follows:

$$EP_j = \hat{\psi}_j - \hat{\phi}_j.$$

Equation 6.1

where $\hat{\phi}$ and $\hat{\psi}$ refer to the seroprevalence estimates before and after standardisation, respectively. Then the *extreme seroprevalence difference* EP is the value of EP_j , for which $|EP_j|$ is at maximum.

In addition, the mean of overall absolute seroprevalence difference or simply *mean seroprevalence difference* MP was defined as:

$$MP = \frac{1}{k} \sum_{j=1}^k |\hat{\psi}_j - \hat{\phi}_j|.$$

Equation 6.2

Note that in order to calculate the mean seroprevalence difference the data were re-grouped into yearly age groups to ensure that all groups were of equal length.

For instance, in the Israel VZV results (Figure 6.4), the extreme seroprevalence difference was observed for the 19 years age group i.e. $EP = -13\%$. The mean seroprevalence difference was $MP = 2\%$.

Tables 6.3 (a)-(f) show the assay methods by antigen used in each laboratory, the local assay cut-off values and the extreme and mean seroprevalence differences following standardisation. In addition, the tables indicate whether the back-standardisation method was used (the back-standardisation method was explained in Section 4.10).

Seroprevalence was only estimated for measles, mumps, rubella, diphtheria, VZV and HAV. Pertussis serosurveys were excluded from the analysis since there are questionable serological correlates of protection. For the purposes of the ESEN2 project, pertussis titres were examined after they were categorised according to their titre measurement

(Giammanco A *et al.*, 2008). HBV serosurveys were also excluded because seroprevalence estimation was more complex, requiring the testing of antibodies against several different antigens (Section 2.4).

Table 6.3 Comparison of standardised and non-standardised seroprevalence estimates by laboratory and assay method used

(a) Measles

country	assay	back-standardisation	local	age group		
	method*	method used	cut-off	(yrs)**	EP (%)	MP (%)
Australia	DB	no	0.15	10	-4.4	0.9
Belgium	HC	no	27	11	71.0	52.1
Bulgaria	HC	no	27	30-39	16.5	9.4
Cyprus	ER	no	0.25	14	26.7	10.2
Czech Republic	SK	no	0.2	1-4	-5.4	2.4
Germany	DB	no	0.15		reference	
Hungary	VR	no	0.15	20-29	-6.7	0.9
Ireland	Biotech	no	0.9	5-9	3.1	2.2
Israel	DB	yes	0.15	8	-2.0	0.4
Latvia	HU	no	0.4	10-19	14.6	5.6
Lithuania	DB	no	0.15	20-24	18.5	3.8
Luxembourg	DB	no	0.15		no difference	
Malta	VR	no	0.15	5-9	0.4	0.2
Romania	DB	no	0.15	11, 14	2.6	1.0
Slovakia	VT	no	9	25-29	-3.6	0.6
Slovenia	DB	no	0.17	20-24	4.5	1.5
Spain	DB	yes	0.15	3	-1.6	0.2
Sweden	In-house	yes	0.12	8	9.5	4.1
UK	DB	no	0.15	11	10.2	1.6

(b) Mumps

country	assay method*	back-standardisation method used	local cut-off	age group (yrs)**	EP (%)	MP (%)
Belgium	HC	no	27	1	17.9	2.9
Bulgaria	HC	no	27	3	-34.1	20.2
Cyprus	ER	no	20	10	22.0	12.5
Czech Republic	SK	no	200	1	-20.0	5.5
Finland	DB	no	230	13	-3.1	1.3
Germany	DB	no	230		reference	
Hungary	VR	no	70	5	2.0	0.4
Ireland	Capita	no	0.9	8	-2.1	0.4
Israel	DB	yes	231	5	-18.6	5.7
Latvia	HU	no	39	1	8.0	5.1
Lithuania	DB	no	150	14	-4.8	1.4
Luxembourg	DB	no	231	13	-9.0	2.7
Malta	VR	no	150	6	22.5	8.9
Romania	DB	no	150	25-29	-5.7	3.3
Slovakia	VT	no	9	7	1.9	0.4
Slovenia	DB	no	190	19	-11.0	2.9
Spain	DB	yes	230	7	-13.0	5.0
Sweden	DB	yes	230	2	-12.8	7.8
UK	DB	no	230	12	-4.0	1.0

(c) Rubella

	assay	back-standardisation	local	age group		
country	method*	method used	cut-off	(yrs)**	EP (%)	MP (%)
Australia	DB	no	5	1	30.2	8.1
Belgium	DO	no	10	9, 10	30.1	16.1
Bulgaria	PL	no	27	5	6.0	1.9
Cyprus	Radim	no	8	14	18.9	5.9
Czech Republic	DO	no	8	10-14, 15-19	2.2	1.3
Germany	DB	no	4		reference	
Hungary	DO	no	10	5-9	7.7	5.1
Ireland	Biokit	no	10	10	3.3	0.8
Israel	DB	yes	4	4	-4.5	0.8
Latvia	PL	no	0.4	1	-18.7	4.5
Lithuania	DB	no	4		no difference	
Luxembourg	DB	no	6		no difference	
Malta	Abbott	no	5	40+	-1.1	1.0
Romania	DB	no	4	3	-1.8	0.4
Slovakia	DO	no	9	7	-1.9	0.4
Slovenia	DB	no	4	14	-1.0	0.3
Sweden	In-house	yes	8		no difference	
UK	Microgen	no	0.89	1	7.7	1.6

(d) Diphtheria

	assay	back-standardisation	local	age group		
country	method*	method used	cut-off	(yrs)**	EP (%)	MP (%)
Australia	DAE	yes	0.01	30-39	3.0	0.8
Czech Republic	NT	no	0.01	50+	-5.9	2.2
Finland	NT	no	0.01		reference	
Hungary	Passive Haem	yes	0.01	50+	-20.8	4.4
Ireland	NT	no	0.016		no difference	
Israel	DAE	no	0.01	60+	-11.6	3.4
Latvia	ELISA (Hycor)	yes	0.01	50+	-8.9	4.3
Luxembourg	NT	no	0.01		no difference	
Slovakia	NT	yes	0.01	20-29	14.5	10.5

(e) VZV

country	assay	back-standardisation	local	age group	EP (%)	MP (%)
	method*	method used	cut-off	(yrs)**		
Belgium	DB	no	50	6, 8, 13	-2.1	0.8
Finland	DB	no	50	8	-4.2	0.2
Germany	DB	yes	50	6	-0.8	0.1
Ireland	DiaMedix	no	15	4	4.3	1.9
Israel	DB	no	50	15-19	-13.0	1.8
Italy	DB	yes	50	2	-2.8	0.6
Luxembourg	DB	no	50	5	-1.4	0.5
Netherlands	HU	no	0.3	10-14	-0.3	0.0
Slovakia	ER	no	250	3	16.2	5.3
Spain	DB	no	50		reference	
UK	DiaMedix	yes	15	1	10.3	5.3

(f) HAV

country	assay	back-standardisation	local	age group	EP (%)	MP (%)
	method*	method used	cut-off	(yrs)**		
Belgium	DiaSorin	no	10	2	20.2	9.4
Czech Republic	Abbott (AxSYM)	no	1	40+	-23.3	15.1
Finland	Enzygost	no	10		no difference	
Germany	Abbott (AxSYM)	yes	1		no difference	
Greece	Abbott (AxSYM)	no	10		reference	
Ireland	DiaSorin	no	10	7	18.1	9.1
Italy	DiaSorin	no	10	4	-29.9	8.0
Lithuania	Enzygost	no	10	20-24	-39.9	17.1
Malta	Abbott (AxSYM)	no	1	7	3.4	0.6
Netherlands	Abbott (AxSYM)	yes	1	9, 25-29	-2.4	1.8
Romania	Abbott (AxSYM)	no	10	1	7.2	0.8
Slovakia	DiaSorin	no	20	2	0.9	0.3

* Assay abbreviations: ER: Euroimmun; DAE: Double Antigen ELISA; DB: Dade Behring; DO: ETI-RUBEK-G PLUS; HC: Hycor; HU: Human; PL: Platelia Rubella IgG; SK: Seiken; VR: Virion; VT: Viotech

** Note that “age group (yrs)” denotes the age group(s) with the highest maximum seroprevalence difference.

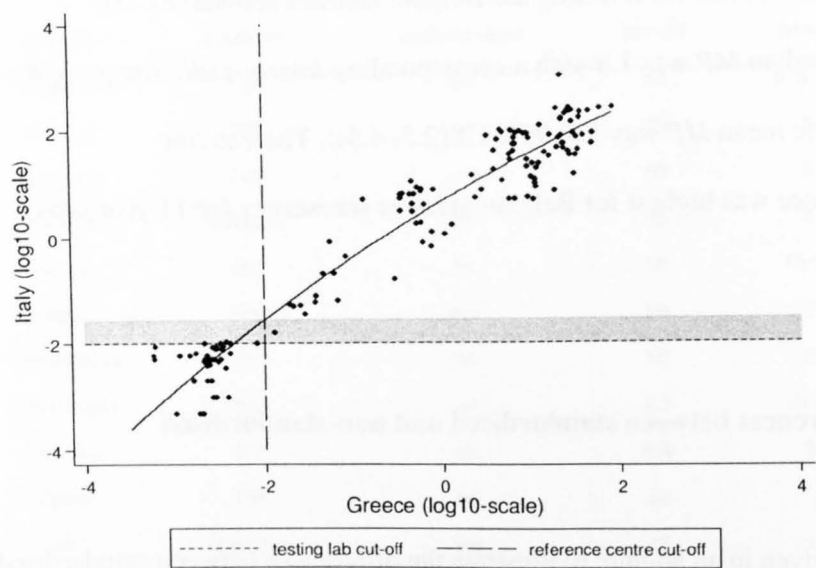
There were a number of laboratories where the mean seroprevalence difference was 0, whereas the highest *MP* resulted from testing the Belgian measles serosurvey (*MP* = 52.1%). The overall median *MP* was 1.6 with a corresponding *interquartile range (IQR)*: [0.4, 5.1] (the arithmetic mean *MP* was 3.5; 95% CI:[2.5, 4.5]). The extreme seroprevalence difference was highest for Belgian measles serosurvey for 11 year-olds ($EP_{11} = 71.0\%$).

6.6 Examples of differences between standardised and non-standardised seroprevalence

Several examples are given in an attempt to illustrate the differences between standardised and non-standardised seroprevalence. The first example is that of the Italian serosurvey results for HAV. As shown in Table 6.3(f), the Italian laboratory used a different assay (DiaSorin) to the reference centre (Abbott (AxSYM)). Although both assays had the same cut-off value (10 IU/ml), a great number of samples were re-classified following standardisation, which meant that both the maximum seroprevalence and the mean absolute seroprevalence differences were large ($EP_4 = -29.9\%$ and $MP = 8.0\%$, respectively).

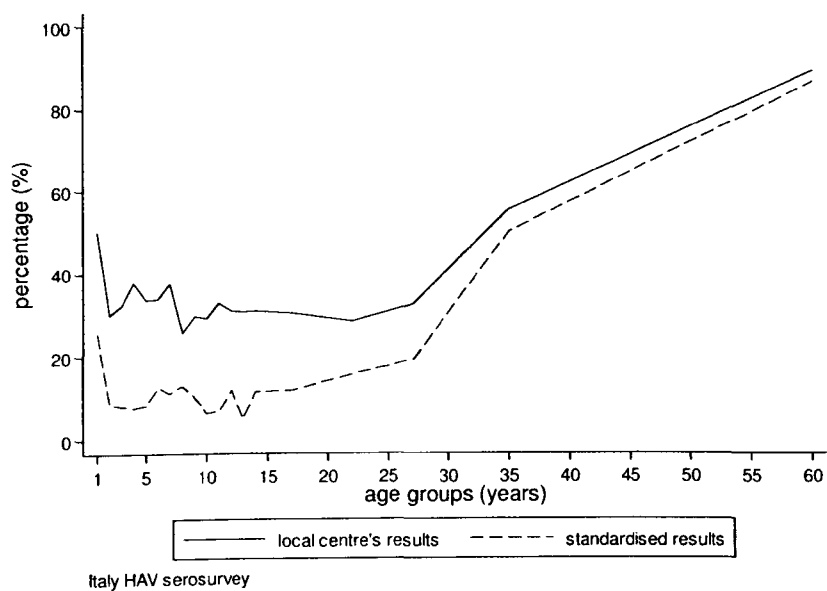
Looking at the standardisation panel results produced by both laboratories, and the standardisation curve, the Italian assay gave higher measurements around the positive / negative cut-off (Figure 6.6). This means that, following standardisation, positive serosurvey samples just over the Italian cut-off (shaded area) were re-classified as negative.

Figure 6.6 Standardisation curve for Italian HAV results



The estimated seroprevalences for Italian serosurvey results tested against HAV before and after these were standardised are shown in Figure 6.7. It seems that there is a high number of low positive results for Italy for the under 25 year-olds, which translates into a seroprevalence difference between standardised and non-standardised results. Therefore, for this example, seroprevalence differences can be attributed to laboratory testing, although it is unclear whether (or how much) the difference is a result of the different assays, or is due to differences between laboratory practices.

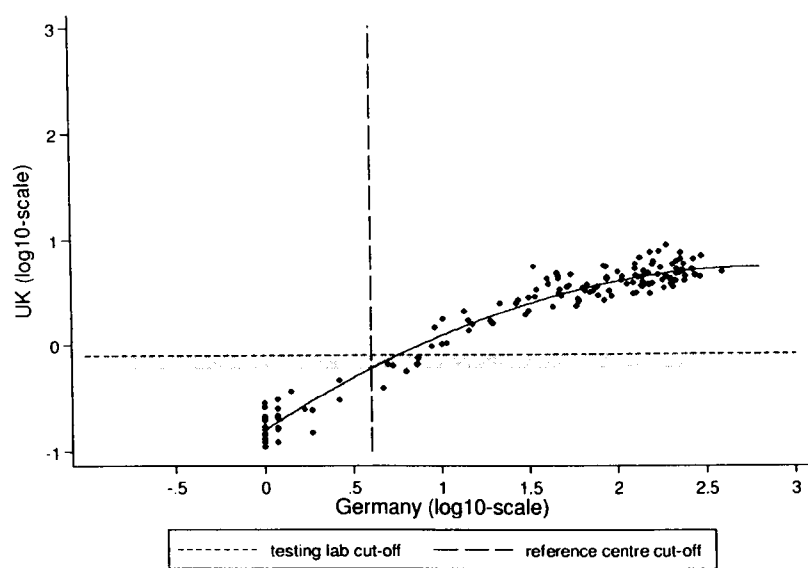
Figure 6.7 Differences between standardised and non-standardised seroprevalence of Italian HAV results by age



The second example is that of the UK rubella serosurvey results for VZV. These were obtained using a different assay (Microgen with an assay cut-off of 0.8 units) to that used for the reference centre's samples (Dade Behring with a cut-off of 4 IU/ml). Although on average the seroprevalence estimates did not differ much after standardisation ($MP = 1.6\%$), there was a high discrepancy between the estimates for the 1 year-olds ($EP_1 = 7.7\%$).

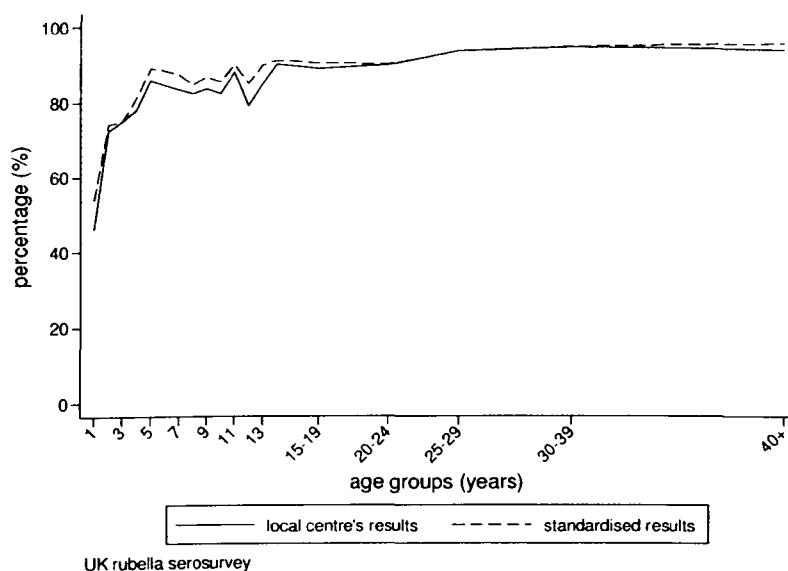
Looking at the standardisation curve, any negative results for the UK that were just below the local assay cut-off (shaded area) were re-classified as positive following standardisation (Figure 6.8).

Figure 6.8 Standardisation curve for UK rubella results



Given that the low negative UK results were re-classified as positives following standardisation, the standardised seroprevalence estimates are higher than the non-standardised (Figure 6.9). This is the opposite effect to the previous example, where the standardised seroprevalence estimates were lower than the non-standardised. Note that in this example, the extreme seroprevalence difference is positive since the standardised estimate is higher.

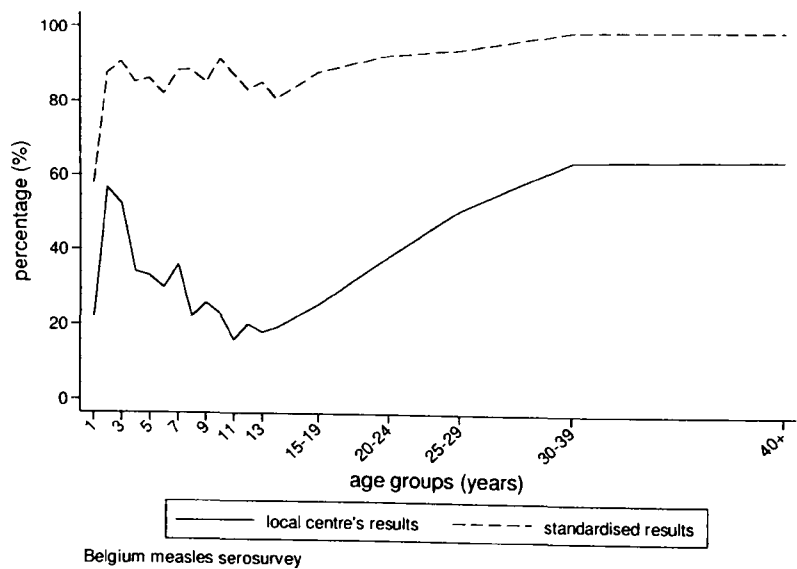
Figure 6.9 Differences between standardised and non-standardised seroprevalence of UK rubella results by age



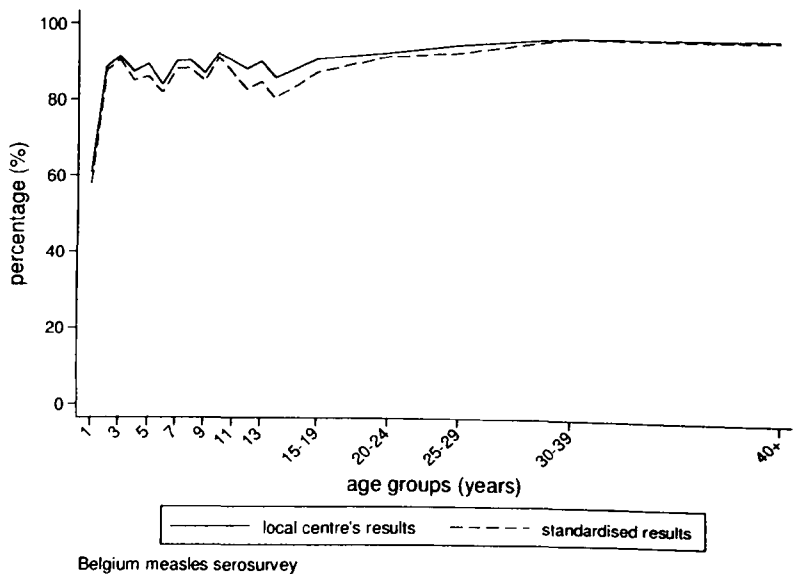
The third example is that of the Belgium serosurvey results for measles. Here there are large differences between standardised and non-standardised seroprevalence ($EP_{11} = 71.0\%$ and $MP = 52.1\%$). Figure 6.10(a) shows that the difference between the two curves follows a similar pattern by age. Figure 6.10(b) shows the same data but multiplied by 10 before taking logs. In this case, standardised and non-standardised seroprevalence match, which raises the question of whether the non-standardised data, tested by the Hycor assay, were reported on the right scale.

Figure 6.10 Differences between standardised and non-standardised seroprevalence of Belgian measles results by age

(a) Original scale for local centre's results



(b) Local centre's results multiplied by 10



6.7 Variability due to different assays

Comparisons of mean seroprevalence differences allow us to investigate different sources of variability. This may occur due to different laboratory techniques or the use of a variety

serological assays and it is important to understand how these varied for individual antigens and how they were affected by age.

Assessing the variability that occurs between different assays was possible by contrasting seroprevalence between laboratories that used the same assay as the reference laboratory, and seroprevalence between laboratories using different assay types. In order to examine variability, the results were further summarised by classifying the mean seroprevalence difference into 3 groups: < 5%, 5% - 10%, and > 10% (Table 6.5).

Table 6.5 Mean seroprevalence difference between laboratories that used the same and different assays as the reference laboratory

assay used	mean seroprevalence difference			Total
	< 5%	5 - 10%	>= 10%	
same assay	32 (84.2%)	4 (10.5%)	2 (5.3%)	38 (100%)
different assay	26 (59.1%)	12 (27.3%)	6 (13.6%)	44 (100%)
Total	58 (70.7%)	16 (19.5%)	8 (9.8%)	82 (100%)

Overall, 84% of the laboratories that used the same assays as the reference laboratories had a mean difference between standardised and non-standardised seroprevalence of less than 5%. A significantly higher proportion of laboratories that used different assays to the reference centres had *MP* higher than 5% ($p = 0.013$). This suggests a smaller standardisation effect on seroprevalence for laboratories that used the same assay as the reference centre.

6.8 Variability arising due to different antigens

As shown before, there is a difference in mean seroprevalence by antigen which could explain part of the overall differences between standardised and non-standardised estimates. Table 6.6 shows that there were on average higher differences between

standardised and non-standardised seroprevalence for HAV ($MP \geq 5\%$ for 45% of the laboratories participating in the serosurvey), or mumps (44% of the participant laboratories had $MP \geq 5\%$) compared to other antigens (less than 25% of the serosurveys had $MP \geq 5\%$).

Table 6.6 Mean seroprevalence difference between antigens

antigen	mean seroprevalence difference			Total
	< 5%	5 - 10%	>= 10%	
measles	14 (77.8%)	2 (11.1%)	2 (11.1%)	18 (100%)
mumps	10 (55.6%)	6 (33.3%)	2 (11.1%)	18 (100%)
rubella	13 (76.5%)	3 (17.6%)	1 (5.9%)	17 (100%)
diphtheria	7 (87.5%)	0 (0%)	1 (12.5%)	8 (100%)
VZV	8 (80%)	2 (20%)	0 (0%)	10 (100%)
HAV	6 (54.5%)	3 (27.3%)	2 (18.2%)	11 (100%)
Total	58 (70.7%)	16 (19.5%)	8 (9.8%)	82 (100%)

6.9 Variability due to different age groups

For the youngest and oldest age groups (1-4 and 20+ year-olds, respectively) the maximum difference between standardised and non-standardised seroprevalence seems to be higher than in the other age groups (approximately 70% of the serosurveys had $EP \geq 5\%$ for these age groups - Table 6.7). The generally higher EP for these age groups are due to a high number of low positive results that were classified differently following the application of the standardisation equation. For those older than 20 years the reason may be waning immunity.

Table 6.7 Extreme seroprevalence difference between age groups

age group	extreme seroprevalence difference			Total
	< 5%	5 - 10%	>= 10%	
1-4	6 (30%)	4 (20%)	10 (50%)	20 (100%)
5-9	12 (63.2%)	3 (15.8%)	4 (21.1%)	19 (100%)
10-19	9 (52.9%)	1 (5.9%)	7 (41.2%)	17 (100%)
20+	4 (28.6%)	4 (28.6%)	6 (42.9%)	14 (100%)
Total	31 (44.3%)	12 (17.1%)	27 (38.6%)	70 (100%)

6.10 Failure of standardisation method

As mentioned above, once the results of a serosurvey have been standardised, the resulting difference between standardised and non-standardised seroprevalence estimates can be attributed to within and between-laboratory variability. However, it may also be a result of a failure of the standardisation method. More specifically, large differences may indicate that laboratory procedures may have changed between the time of testing the standardisation panel and the serosurvey samples. Alternatively, it could be a result of an atypical set of samples used for the standardisation panel, which could have resulted in a biased standardisation equation.

Theoretically, such potential failures should not occur when the back-standardisation method is being used since the panel is chosen from already-tested serosurvey (Section 4.10). In practice, however, a different type of bias would arise from using back-standardisation. This would be due to the deterioration of serum samples with time as the same serosurveys had sometimes been stored for more than 5 years before the beginning of the ESEN2 project.

A comparison between standardised and non-standardised estimated seroprevalence obtained with and without back-standardisation is shown in Table 6.8. There was little difference between the two methods, with standardisation reporting a slightly higher

proportion of serosurveys with *MP* >10% (9% and 6% for standardisation and back-standardisation, respectively), but slightly lower for *MP* between 5% and 10% (19% and 24% for standardisation and back-standardisation, respectively).

Table 6.8 Mean seroprevalence difference between standardisation and back-standardisation methods

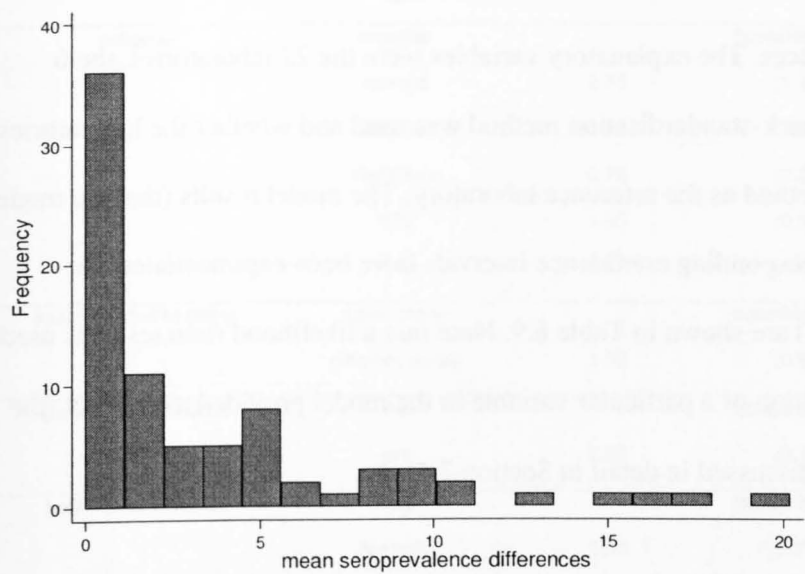
method	mean seroprevalence difference			Total
	< 5%	5 - 10%	>= 10%	
standardisation	46 (71.9%)	12 (18.8%)	6 (9.4%)	64 (100%)
back-standardisation	12 (70.6%)	4 (23.5%)	1 (5.9%)	17 (100%)
Total	58 (71.6%)	16 (19.8%)	7 (8.6%)	81 (100%)

6.11 Multivariable analysis of seroprevalence differences

The tables above show a descriptive interpretation of the seroprevalence difference between the standardised and non-standardised results. The mean seroprevalence difference was also analysed using a multivariable regression model aimed at identifying the factors associated with large differences in seroprevalence estimates.

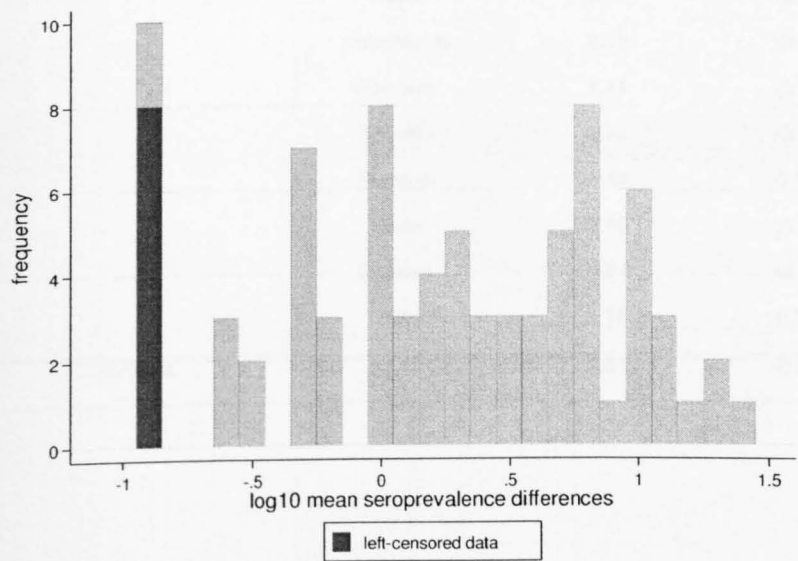
Each of the 22 national laboratories participating in ESEN2 was involved in testing samples of between 2 and 6 antigens. Apart from the reference laboratories where no standardisation was necessary, the Belgian measles data were also excluded, since there was a question of integrity of the data (Section 6.6). The distribution of the mean seroprevalence differences for the remaining 81 serosurveys is shown in Figure 6.11.

Figure 6.11 Distribution of the mean seroprevalence differences



In order to satisfy the normality assumptions the data were \log_{10} -transformed. For 8 serosurveys there was no difference between standardised and non-standardised results at any age group. These were replaced by the lowest mean seroprevalence difference ($MP = \log_{10}(0.1)$), and were treated in the analysis as censored (Figure 6.12).

Figure 6.12 Distribution of the \log_{10} -transformed mean absolute differences



A censored normal regression model was fitted to the log-transformed mean seroprevalence differences. The explanatory variables were the 22 laboratories, the 6 antigens, whether the back-standardisation method was used and whether the laboratories used the same assay method as the reference laboratory. The model results (that the model estimates and their corresponding confidence intervals have been exponentiated for interpretation purposes) are shown in Table 6.9. Note that a likelihood ratio test was used to test whether the addition of a particular variable in the model provided a better fit (the likelihood ratio test is discussed in detail in Section 7.4).

Table 6.9 Multivariable analysis of mean absolute seroprevalence differences

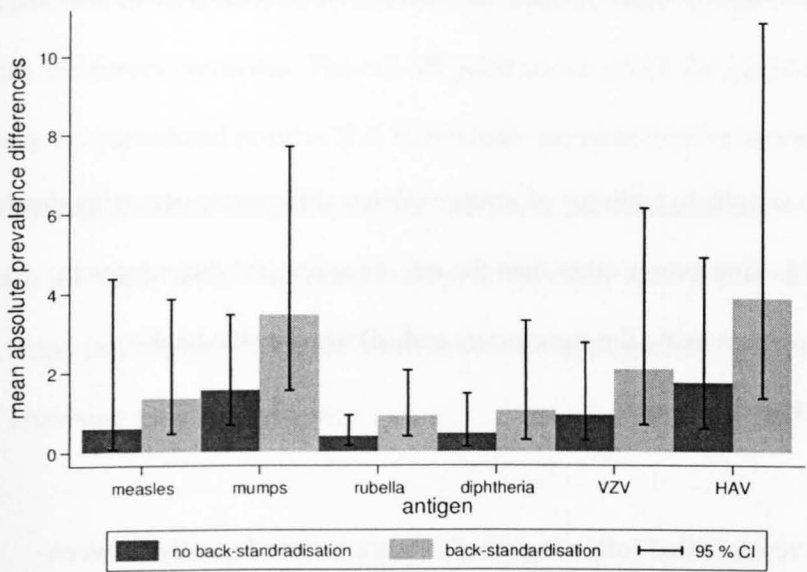
Variable	Level	MP estimates	95% CI	p-value
antigens	measles		baseline	
	mumps	2.58	(1.17, 5.72)	
	rubella	0.68	(0.30, 1.54)	
	diphtheria	0.78	(0.24, 2.47)	
	VZV	1.55	(0.53, 4.56)	
	HAV	2.87	(1.02, 8.07)	0.014
variability due to assay	same assay		baseline	
	different assay	1.32	(0.62, 2.82)	0.477
back-standardisation	no		baseline	
	yes	2.22	(0.77, 6.35)	0.140
national laboratories	Italy		baseline	
	Australia	2.57	(0.27, 24.84)	
	Belgium	3.46	(0.43, 28.18)	
	Bulgaria	7.37	(0.73, 73.89)	
	Cyprus	9.41	(0.94, 94.40)	
	Czech Republic	3.78	(0.46, 30.95)	
	Finland	0.16	(0.02, 1.49)	
	Germany	0.02	(0.00, 0.26)	
	Hungary	1.60	(0.18, 14.22)	
	Ireland	0.91	(0.12, 7.00)	
	Israel	1.17	(0.15, 8.94)	
	Latvia	4.59	(0.52, 40.88)	
	Lithuania	1.57	(0.18, 13.65)	
	Luxembourg	0.22	(0.02, 1.97)	
	Malta	0.91	(0.11, 7.70)	
	Netherlands	0.13	(0.01, 1.45)	
	Romania	1.11	(0.13, 9.90)	
	Slovakia	0.92	(0.12, 6.76)	
	Slovenia	1.49	(0.15, 15.31)	
	Spain	0.46	(0.04, 5.71)	
	Sweden	0.64	(0.06, 6.59)	
	UK	1.76	(0.22, 14.17)	0.003
constant		0.61	(0.08, 4.37)	0.622

Antigens were significantly associated with mean seroprevalence difference ($p = 0.014$) and in particular, mumps and HAV had the highest *MP* estimates. There were significant differences between *MP* by different laboratories ($p = 0.003$). The Bulgarian and Cypriot laboratories had the highest *MP* estimates which were significantly higher than the German laboratory, where the smallest differences were observed.

Using a different assay from the reference laboratory did not affect the number of samples re-classified following standardisation, and hence, the mean seroprevalence difference ($p = 0.48$). Using back-standardisation seemed to generate higher mean seroprevalence differences, however, this finding was not significant ($p = 0.14$).

The mean seroprevalence difference estimates for different antigens, and whether back-standardisation was used, are also shown in Figure 6.13 below. The back-standardisation method seems to consistently give higher seroprevalence differences although the confidence intervals overlap.

Figure 6.13 Multivariable model results of mean absolute seroprevalence differences by antigen and whether back-standardisation was used (adjusted for laboratories and whether labs used the reference assay)



As an alternative analysis method, the mean seroprevalence differences were grouped into “<5%” and “≥5%”. A logistic regression model was fitted using the same explanatory variables as in Table 6.8. There was significant association between the mean seroprevalence difference and antigens ($p = 0.009$) with mumps and HAV having higher odds of mean seroprevalence difference of 5% or higher. There was no evidence of an association between more than 5% seroprevalence differences and testing in different laboratories ($p = 0.16$), using the back-standardisation method ($p = 0.18$) or when using different assays than the reference centre ($p = 0.62$). Note that different MP classifications gave slightly different model estimates; there were generally significant differences between antigens and borderline significant evidence for different laboratories.

6.12 Conclusions

The main aim of this Chapter was to quantify the impact of standardisation on seroprevalence estimation. Overall, the median change of seroprevalence estimates following standardisation was 1.6 (IQR: [0.4, 5.1]) whereas the geometric mean was 3.5 (95% CI: [2.5, 4.5]).

A secondary aim was to test the hypotheses of seroprevalence differences occurring due to: (a) different antigens, (b) using assays other than the reference one, (c) differences in laboratory methods even when using the same assay and (d) whether the back-standardisation method had been used.

The number of samples re-classified following standardisation varied greatly between antigens, with HAV and mumps showing the highest change on seroprevalence estimates. For both antigens, many serological assays performed poorly with high variability around the standardisation equations which meant that there was an increased risk for samples to be misclassified. Moreover, the changes in seroprevalence can be explained by between-laboratory variability, with a high number of samples re-classified following standardisation for the Cypriot and the Bulgarian laboratories.

As mentioned above, in theory, back-standardisation might provide a better technique because it ensures there are no differences between the standardisation panel and the serosurvey serological samples. This was not shown from the analysis results, where the sample's re-classification following standardisation was higher using back-standardisation (although this finding was not significant). The reason for this may be a different type of bias attributed to the deterioration of the serum samples with time. In the case of back-standardisation it may be several years between the time the serosurvey is tested and the time the standardisation panel is chosen and tested. Hence, there is little benefit from using

back-standardisation, and therefore, it is only advised to be used as an alternative method to the standard method that was most commonly used during the ESEN2 project.

In addition to variability and potential bias, another issue appears important when dealing with serosurvey estimates. The cut-off point above which the quantitative results of an assay are considered positive (i.e. individuals are protected) is currently defined by the assay manufacturer using a variety of methods. The choice of this cut-off is very important since it directly affects the classification of the assay results. This issue will be addressed in detail in Chapter 8. Chapter 7 will introduce mixture modelling as an alternative method of analysing serological results.

Chapter 7: Mixture models as an alternative method to standardisation

7.1 Introduction - A standardisation limitation

As discussed in the previous Chapters, once the serosurvey results had been standardised to common units, these were classified as positive and negative in order to estimate the population seroprevalence. This classification was carried out in accordance to a cut-off point specified by the reference centre's assay manufacturer. The choice of this cut-off is very important since it directly affects the classification of the assay results of all participant laboratories.

A detailed discussion on how these cut-offs were chosen by the assay manufacturers and alternative techniques for estimating such cut-offs, is provided in Chapter 8. The present Chapter will address the question of whether the method of standardisation and the definition of cut-offs can be bypassed altogether by the application of *mixture models* directly on the non-standardised serological results. Several types of mixture models will be presented, and a comparison between seroprevalences obtained by standardisation and mixture modelling will be carried out. The advantages and disadvantages of each method will be discussed.

7.2 Background on mixture modelling

The standardisation method was developed to compare seroprevalence estimates for the ESEN2 project. An alternative method for seroprevalence estimation that has been used in the past is mixture modelling, which has been described as “a statistical method that can be used to reveal component distributions in a heterogeneous population” (Dong Z, 1997).

Mixture model methodology has been developing for the last 100 years, however due mainly to computational difficulties, it is only since the 1980s that considerable advances have been made in fitting these models, and in particular using the method of maximum likelihood (McLachlan G & Peel D, 2000). As an illustration of these difficulties, in 1894

Karl Pearson had to solve a ninth degree polynomial using the method of “moments estimators” for a mixture model with two normal components. Since the development of modern computers, mixture models have become popular in many scientific areas using maximum likelihood to obtain parameter estimates (Dong Z, 1997; McLachlan G & Peel D, 2000).

In terms of applications to serological data, a mixture model was first used in 1990 for estimating an assay cut-off that accurately classifies individuals’ sera (Parker RA *et al.*, 1990). That particular application of mixture modeling will be discussed in detail in Chapter 8. A mixture model was first used to estimate population seroprevalence directly from the data, rather than initially classifying individual samples into “positive” or “negative”, in a parvovirus serosurvey carried out in 1991 (Gay NJ, 1996). Since then, amongst other studies, they have been used to model VZV (Vyse AJ *et al.*, 2004) and measles, mumps and rubella data (Vyse AJ *et al.*, 2006; Hardelid P, 2008).

The mixture modelling method is presented below as it may be generally applied. This will allow for comparison with the standardisation algorithm to illustrate the major advantages and drawbacks of this methodology.

7.3 A simple mixture model

A generalised equation for the probability density function of a mixture model with components from the same family of distributions can be defined as:

$$g(x) = \sum_j \pi_j g(x; \theta_j),$$

Equation 7.1

where π_j is the probability that an observation comes from component $j = 1, \dots, K$

$\left(\sum_j \pi_j = 1 \right)$. Within component j , x has density $g(x; \theta_j)$ where θ_j are parameters that vary

between components (McLachlan G & Peel D, 2000).

Since the serosurveys examined were age stratified, the mixture models used should be estimated within age groups, which means that the model in Equation 7.1 becomes:

$$g_i(x) = \sum_j \pi_{ij} g(x; \theta_{ij}),$$

Equation 7.2

for $i = 1, \dots, \zeta$ age groups.

A simple mixture model is applicable to serological survey data assuming:

1. Two underlying distributions corresponding to two distinct population groups, one for negative samples / susceptibles and one for positive / protected individuals. Let $j = 0$ denote the negatives and $j = 1$ the positives, and set $\pi_{i1} = \pi_i$, $\pi_{i0} = 1 - \pi_i$. Hence, π_i is the seroprevalence for age group i .

2. The underlying component distributions are normal.

3. The shapes of the underlying distributions (location and dispersion parameters) are common across age groups and therefore, do not change with age.

Given these assumptions, Equation 7.2 becomes:

$$g_i(x) = (1 - \pi_i) f_N(x; \mu_-, \sigma_-) + \pi_i f_N(x; \mu_+, \sigma_+),$$

Equation 7.3

where x denotes the \log_{10} -transformed IU/ml, $f_N(x; \mu_-, \sigma_-)$, $f_N(x; \mu_+, \sigma_+)$ are the normal density functions for the negative and positive populations, respectively and π_i stands for seroprevalence in age group i ,

The model parameters μ_- , σ_- , μ_+ , σ_+ and π_i of the Equation 7.3 can be estimated using maximum likelihood. The log-likelihood function used for such maximisation is obtained as a sum of terms:

$$\ln(L_{hi}) = \ln(g_i(x_h)),$$

Equation 7.4

where x_h , $h = 1, \dots, n_i$ are the quantitative observations in age group i .

For data below the assay detection range (left-censored), the log-likelihood becomes:

$$\ln(L_{ci}) = \ln\left[(1 - \pi_i) F_N(x_c; \mu_-, \sigma_-) + \pi_i F_N(x_c; \mu_+, \sigma_+)\right],$$

Equation 7.5

where F_N is the standard normal cumulative distribution function and x_c , $c = 1, \dots, \ell_i$ are the left-censored observations in age group i .

Similarly, for right-censored data, the log-likelihood is:

$$\ln(L_{di}) = \ln \left[1 - (1 - \pi_i) F_N(x_d; \mu_-, \sigma_-) - \pi_i F_N(x_d; \mu_+, \sigma_+) \right],$$

Equation 7.6

where $x_d, d = 1, \dots, m_i$ are the right-censored observations in age group i .

7.4 Model diagnostics

Once a mixture model is fitted to the data, it is important to look at model fit and investigate whether model assumptions hold. There are several diagnostic criteria that can be used (a) to compare different models which are either nested or not and (b) to assess the overall fit of a model. Some of these are listed below.

AIC and BIC

One method for making comparisons between models is the *Akaike Information Criterion (AIC)* defined as

$$AIC = -2 \ln(L) + 2k,$$

Equation 7.7

where $\ln(L)$ is the maximised log-likelihood of the model, and k denotes the number of parameters in the model.

An alternative method is the *Bayesian Information Criterion (BIC)* defined as

$$BIC = -2 \ln(L) + k \ln(n),$$

Equation 7.8

where n is the number of observations.

In each case, the model with the lowest value of the chosen information criterion is selected.

These approaches penalise models with high number of parameters, with BIC producing a greater penalty as compared to AIC (Morgan BJT, 2000).

Likelihood ratio test

AIC and BIC are often used to make comparisons between models since they are easy to derive. However, when the models to be compared are nested within each other, i.e. the more complex ones can be transformed into the simpler ones by applying a set of linear constraints on the parameters, then the *likelihood ratio test* or *deviance test* can be used instead.

Suppose L_1 is the maximum likelihood of model 1 and L_2 the maximum likelihood of model 2 that is nested within model 1 i.e. model 2 is a simplification of model 1. For testing the null hypothesis that model 2 is the true model, the likelihood ratio test statistic D is defined as:

$$D_{df_2 - df_1} = -2 \ln \left(\frac{L_2}{L_1} \right) = -2 (\ln(L_2) - \ln(L_1)) .$$

Equation 7.9

Under the null hypothesis that model 2 is the true model, the statistic D follows an approximate chi-square distribution with $df_2 - df_1$ degrees of freedom, where df_1 , df_2 are the

degrees of freedom for the model 1 and model 2, respectively (Collett D, 1999; Kuha J, 2004).

Pearson Goodness-of-fit test

There are different methods of testing the model fit, the most commonly used being the *Pearson Goodness-of-fit test* (Kirkwood BR & Sterne JAC, 2003; Morgan BJT, 2000). An often quoted disadvantage of this method is that the data need to be grouped into bins prior to carrying out the test, and therefore, the results depend on how these bins have been set. However, having to define the bins may also be an advantage since it makes it possible to focus around the area of interest e.g. around the region where densities meet for the mixture models presented here (and hence, where misclassification is most likely to occur).

To test the null hypothesis that the data follow a specific distribution, assume the data have been divided into N bins. For a bin i defined by the interval $[X_L, X_U]$ the expected frequency is:

$$E_i = n(F(X_U) - F(X_L)),$$

Equation 7.10

where n is the total number of observations and F is the cumulative distribution function.

The Pearson test statistic is then calculated by:

$$X^2 = \sum_{i=1}^N \frac{(O_i - E_i)^2}{E_i},$$

Equation 7.11

where O_i is the observed frequency. Under the null hypothesis that the model is correct, the test statistic follows an approximate chi-square distribution with $N - m - 1$ degrees of freedom where m is the number of estimated parameters.

To show where the greatest differences between observed data and fitted distribution lie, the Pearson residuals can be calculated as

$$e_i = \frac{O_i - E_i}{\sqrt{E_i}} .$$

Equation 7.12

Under the null hypothesis we would expect most of the residuals to fall within the range ± 2 . The expected frequencies in each bin need to be sufficiently large (at least 5) for the chi-squared distribution to be correct, and hence for the test to be valid (Zar JH, 1999).

Bootstrapping

Another method used to assess the model fit at different points of the serological distribution is *bootstrapping*. The method is described as follows:

1. Using the observed data y_i ($i = 1, \dots, n$), a random sample \hat{y}_i^* can be drawn using sampling with replacement (bsample command in Stata). This is possible if each measurement selected randomly from n observations, is returned back to the sample before the next measurement is drawn. In other words, each measurement has a probability of $1/n$ of being selected in the first place and can be included in the sample multiple times. This process can be repeated until M random samples of size n have been drawn.

2. Mixture models can be fitted to each of these samples obtaining estimates $\hat{\theta}_{hj}^*$ where $h = 1, \dots, k$ indexes the model parameters and $j = 1, \dots, M$ the bootstrap sample to which the model is fitted.

3. The bootstrap estimate for the h^{th} parameter is then $\frac{1}{M} \sum_{j=1}^M \hat{\theta}_{hj}^*$.

Finally, upper and lower percentile limits around the parameter estimates and other quantities, such as the density, can be presented. Using these bootstrap confidence intervals, it is possible to test the null hypothesis that the model is correct.

Kolmogorov-Smirnov Goodness-of-fit test

An alternative test for continuous data that does not require the data to be binned is the *Kolmogorov-Smirnov Goodness-of-fit test* (Zar JH, 1999). The null hypothesis under this test is that the data are drawn from the fitted distribution. The test is based on the distance between the *empirical cumulative distribution function (ECDF)* and the fitted cumulative distribution function.

Let observations x_i ($i = 1, \dots, n$) be ordered from smallest to largest. Then the ECDF can be defined as:

$$E_n(x_i) = \frac{n(i)}{n},$$

Equation 7.13

where $n(i)$ are the number of points less than x_i .

The Kolmogorov-Smirnov statistic D_n is then defined as the maximum of the distances

$$D_n = \max |E_n(x_i) - F(x_i)|,$$

Equation 7.14

where $F(x_i)$ is the cumulative distribution function.

The test statistic D can be compared to the Kolmogorov-Smirnov test critical values $D_{\alpha,n}$ where α is the significance level. Note that the Kolmogorov-Smirnov test is more sensitive near the centre of the distribution than its tails (Wilcox R, 1998).

7.5 Example of a mixture model (naïve model)

The Belgium VZV serosurvey results are used to demonstrate the comparison between standardisation and mixture model methods. VZV was chosen rather than other antigens included in ESEN2 because no routine vaccination schedule exists. This greatly simplifies the mixture model since it can be assumed that there are only two underlying populations (individuals who had either previously been infected or never been infected).

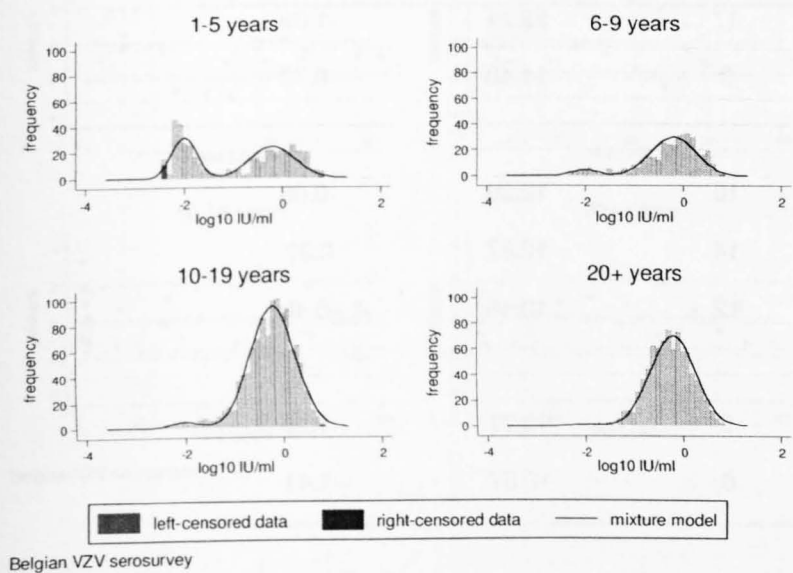
A mixture model was fitted to the 2,762 standardised results given by the Belgian laboratory. Thirteen samples of extremely low IU/ml results were classified as “censored” and were fixed at 0.003 IU/ml (the lowest value in the dataset) or at -2.5 on the \log_{10} -scale. Similarly 3 sera of extremely high results were fixed at 5.8 IU/ml (or at 0.8 on the \log_{10} -scale).

As in Equation 7.3, two underlying populations (for negative/ non-infected and positive / infected individuals) were assumed, both following normal distributions with common

location and dispersion parameters by age group. The Stata program for fitting such a model is given in Appendix I(C).

The mixture distribution by age group following model estimation (negative component: $\hat{\mu}_- = -2.05$, $\hat{\sigma}_- = 0.25$; positive component: $\hat{\mu}_+ = -0.26$, $\hat{\sigma}_+ = 0.45$; prevalence estimates: $\hat{\pi}_1 = 0.58$, $\hat{\pi}_2 = 0.92$, $\hat{\pi}_3 = 0.97$, $\hat{\pi}_4 = 0.99$) is given in Figure 7.1 below. The mixture model appears to not quite fit the observed data within the \log_{10} IU/ml intervals $(-0.5, 0)$ and $(0, 0.5)$, for age groups 1-5 and 6-9 years.

Figure 7.1 Fitted model assuming constant mean and standard deviation for Belgian serosurvey tested against VZV



7.6 Example of model diagnostics (naïve model)

Pearson Goodness-of-fit test

To assess the Goodness-of-fit of the model described in the previous Section, a Pearson chi-square test was initially carried out. As mentioned earlier, this test depends on the

definition of the bins and it requires sufficiently large expected frequencies in each bin in order to be valid. For these reasons, a procedure was defined prior to the analysis.

Initially, the serological results were grouped into 220 bins of different widths. Table 7.1 below shows some of these bins (\log_{10} -transformed data), the observed and the expected counts. Note that the expected values were estimated by age group, and therefore, the data were pooled over the 4 age groups. The Pearson residuals estimated as shown in Equation 7.12 are also shown. These bins were used as long as the pooled expected frequencies were approximately between 10 and 15.

Table 7.1 Serological outcome grouped in 220 bins for Pearson Goodness-of-fit test

interval	observed	expected	residuals
[-2.50, -2.40)	17	22.11	-1.09
[-2.40, -2.34)	9	11.48	-0.73
...
[-0.90, -0.89)	10	12.29	-0.65
[-0.89, -0.87)	14	12.87	0.32
[-0.87, -0.86)	12	13.46	-0.40
...
[0.62, 0.67)	9	13.77	-1.29
[0.67, 0.71)	6	10.57	-1.41

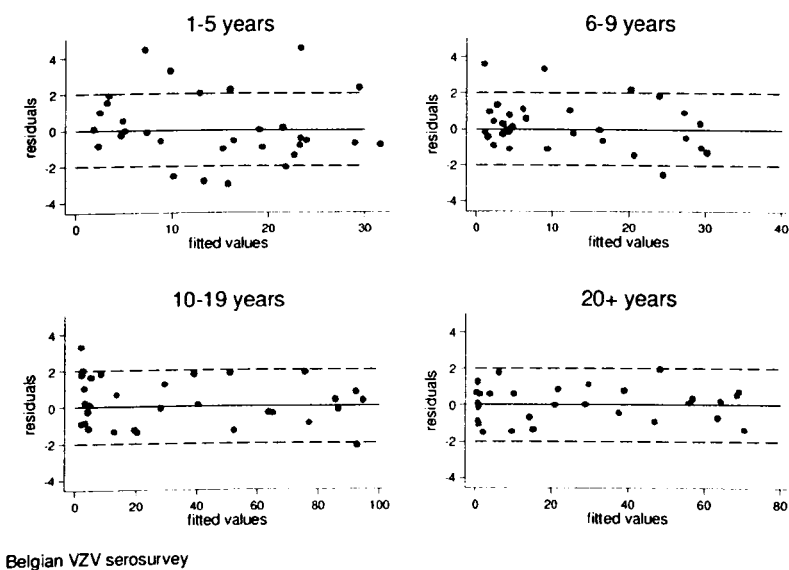
Using Equation 7.11 the test statistic was $X^2 = 276.3$ for 211 degrees of freedom ($p = 0.002$). The test provides evidence to reject the null hypothesis that the model is correct.

By plotting the residuals against their corresponding fitted values it is possible to see whether there is a systematic pattern; such a pattern suggest a bad model fit. Figure 7.2(a) suggests that this is not the case and the residuals are randomly scattered, though there are some residuals outside the ± 2 limits.

To find where the major differences between the observed data and the fitted mixture model lie, the Pearson residuals were plotted against log IU/ml. The critical area where it is important for the model to provide a good fit is around the area where the two underlying components meet. For this example, this is between -1.5 and 0.5 on the \log_{10} -scale. Figure 7.2(b) shows that there is generally a good fit in this area, despite the fact that there is the odd large residual in the age groups 6-9 and 10-19 years.

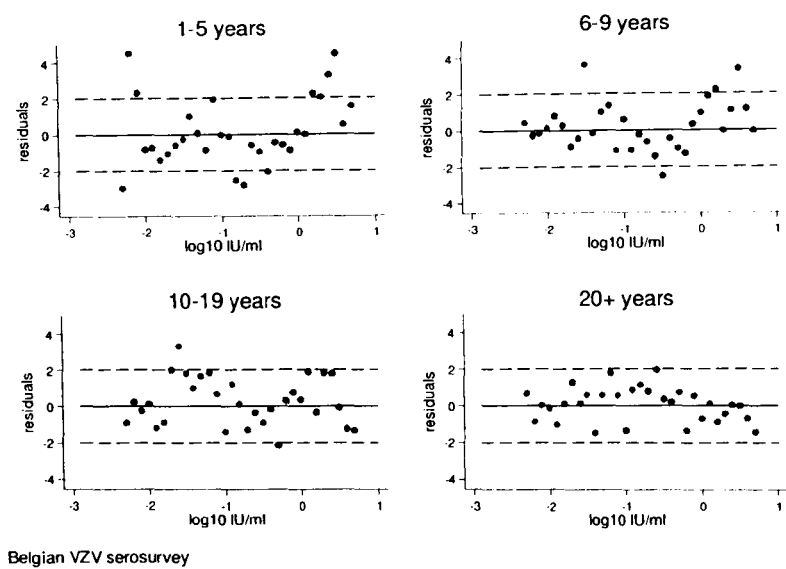
Figure 7.2 Pearson residuals by age group for the naïve model (excl. censored data)

(a) Residuals vs. fitted values



Belgian VZV serosurvey

(b) Residuals vs. log₁₀ IU/ml



The test was repeated after the data were grouped in 30 bins, concentrating on the area where the underlying distributions meet i.e. -2 to -0.8 on the log₁₀-scale. Serological results within the area of interest were grouped into bins of different widths, as long as the pooled expected frequency was between 10 and 20.

Table 7.2 shows the grouped serological results (log₁₀-transformed data) pooled over the 4 age groups. High residuals were reported between [-1.46, -1.33) and [-1.25, -1.15). The Pearson Goodness-of-fit test showed significant differences between observed and expected frequencies ($X^2 = 35.63$; $df = 21$; $p = 0.024$).

Table 7.2 Serological outcome grouped in 30 bins for Pearson Goodness-of-fit test

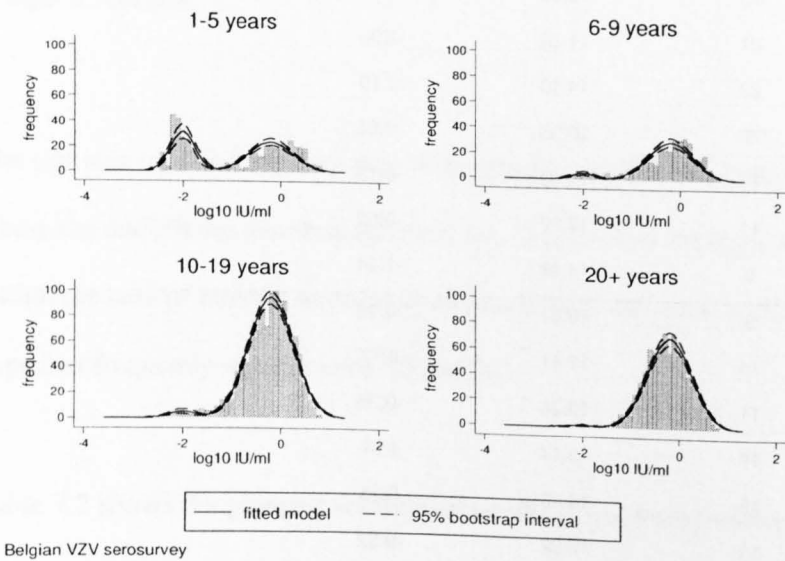
interval	observed	expected	residuals
[-2.50, -2.00)	167	155.96	0.88
[-2.00, -1.97)	11	14.30	-0.87
[-1.97, -1.93)	13	13.62	-0.17
[-1.93, -1.90)	9	12.73	-1.05
[-1.90, -1.86)	10	11.68	-0.49
[-1.86, -1.83)	9	10.53	-0.47
[-1.83, -1.75)	15	18.53	-0.82
[-1.75, -1.69)	10	11.12	-0.33
[-1.69, -1.59)	16	12.92	0.86
[-1.59, -1.46)	16	11.83	1.21
[-1.46, -1.33)	23	14.12	2.36
[-1.33, -1.25)	15	13.38	0.44
[-1.25, -1.20)	21	11.26	2.90
[-1.20, -1.15)	22	14.10	2.10
[-1.15, -1.12)	8	10.05	-0.65
[-1.12, -1.09)	11	11.40	-0.12
[-1.09, -1.06)	11	12.87	-0.52
[-1.06, -1.03)	9	14.48	-1.44
[-1.03, -1.01)	9	10.61	-0.49
[-1.01, -0.99)	14	11.41	0.77
[-0.99, -0.97)	11	12.26	-0.36
[-0.97, -0.95)	19	13.14	1.61
[-0.95, -0.93)	15	14.06	0.25
[-0.93, -0.91)	13	15.02	-0.52
[-0.91, -0.89)	10	16.01	-1.50
[-0.89, -0.87)	21	17.03	0.96
[-0.87, -0.85)	17	18.08	-0.25
[-0.85, -0.83)	20	19.16	0.19
[-0.83, -0.81)	16	20.26	-0.95
[-0.81, 0.80)	2313	2315.13	-0.04

Bootstrapping

As an alternative method to the Pearson residual plot, bootstrapping was used to indicate where the differences between observations and fitted values lie. One thousand samples of size 2,762 were randomly drawn, with replacement from the original dataset and the 2.5th

and 97.5th centiles were calculated for each age group. Figure 7.3 shows the fitted densities together with the 95% bootstrap interval. The fact that many observations lie outside the intervals around the fitted line, underline the lack of good model fit. Although a bootstrap confidence interval can easily be produced for a simple mixture model, this becomes more difficult for complicated models where the initial parameter values need to be carefully chosen for the model to converge correctly.

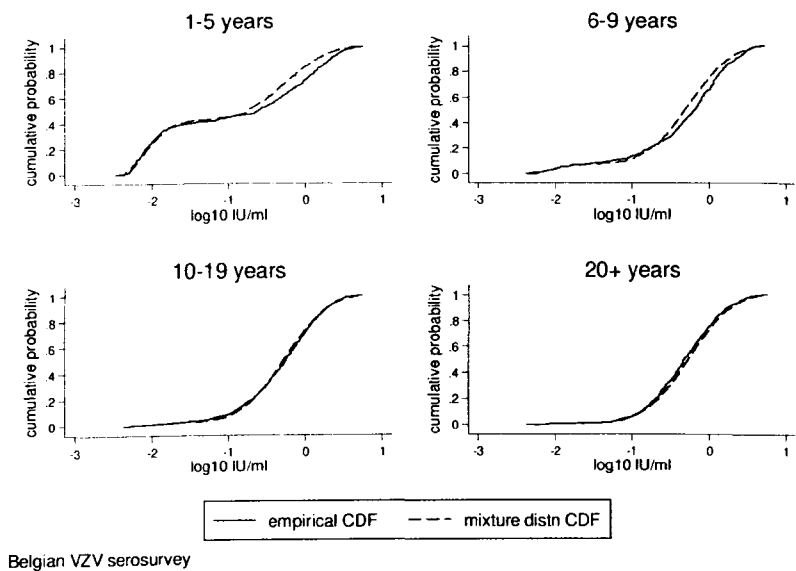
Figure 7.3 Fitted model and observed data with 95% bootstrap intervals for the naïve model



Kolmogorov-Smirnov Goodness-of-fit test

The Kolmogorov-Smirnov Goodness-of-fit test was also used as an alternative to the Pearson chi-square test. After excluding all censored data, the test showed significant differences between the ECDF and the CDF of the fitted model. Using Equation 7.14 the test statistic was calculated as $D_{0.05,2746} = 0.19$ ($p < 0.001$). Figure 7.4 shows that the large differences between the observed and expected cumulative distributions occurred between -0.5 and 0.3 for age groups 1-5 and 6-9 years.

Figure 7.4 ECDF plotted against mixture distribution CDF by age group



7.7 Varying mean and standard deviation by age group

One of the assumptions of the model fitted above is that the parameters of the negative and positive components of the mixture distribution remain constant by age group. Specifically for VZV, very little information is known regarding the tendency of antibody levels to wane over time (Vyse AJ *et al.*, 2004). However, as it has been noted in the past for parvovirus B19 serology, and might be applicable more generally to other vaccine-preventable diseases, “the model for the density of results from seropositives... needs to be sufficiently flexible to describe any decay in antibody levels” (Gay NJ, 1996).

This may be achieved by allowing the shape of the underlying distributions to vary by age. In this case, the shape of the negative component is assumed to be fixed by age, however, the location and dispersion parameters of the positive components may vary to allow for potential waning immunity. Given these assumptions, a model of the following form was fitted:

$$g_i(x) = (1 - \pi_i) f_N(x; \mu_-, \sigma_-) + \pi_i f_N(x; \mu_{+i}, \sigma_{+i}),$$

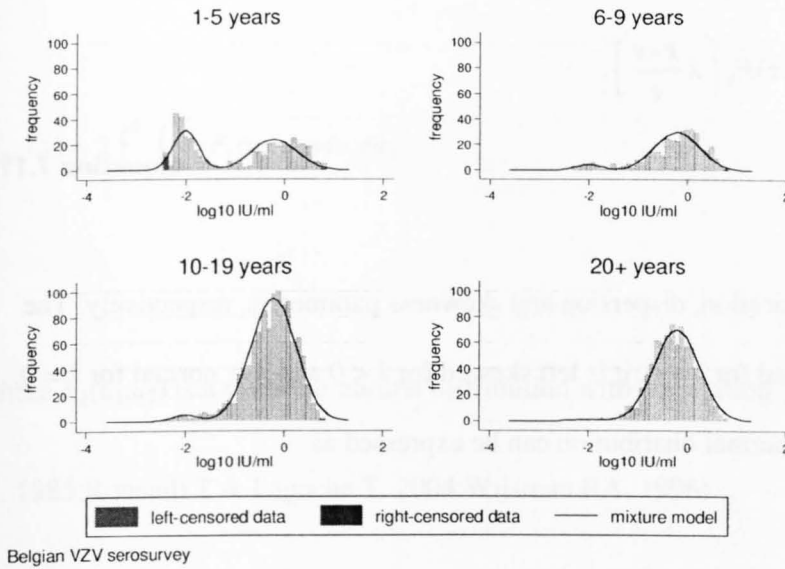
Equation 7.15

where μ_{+i} and σ_{+i} are the mean and standard deviation, respectively, of the positive component for age group i .

For the Belgium VZV serosurvey example, the data can be grouped into 4 age groups ($i = 1, \dots, 4$) i.e. 1-5, 6-9, 10-19 and older than 20 years. This resulted in a 14-parameter-model (negative component: $\hat{\mu}_- = -2.05$, $\hat{\sigma}_- = 0.26$; positive component: $\hat{\mu}_{+1} = -0.11$, $\hat{\mu}_{+2} = -0.19$, $\hat{\mu}_{+3} = -0.27$, $\hat{\mu}_{+4} = -0.31$, $\hat{\sigma}_{+1} = 0.46$, $\hat{\sigma}_{+2} = 0.46$, $\hat{\sigma}_{+3} = 0.45$, $\hat{\sigma}_{+4} = 0.43$; prevalence estimates: $\hat{\pi}_1 = 0.57$, $\hat{\pi}_2 = 0.92$, $\hat{\pi}_3 = 0.97$, $\hat{\pi}_4 = 0.99$).

The fitted model is shown in Figure 7.5 below. Although the model seems to fit the data better when varying the positive distribution component parameters by age group, it may not accurately incorporate the shape of the distribution. There are differences between the observed and fitted distributions between the \log_{10} IU/ml interval $(-0.5, 0.5)$ for age groups 1-5 and 6-9 years and $(0, 0.5)$ for 10-19 years. In particular, for the age groups 1-5 and 6-9 years, the positive component of the frequency distribution seems to be slightly skewed to the left. Moreover, there is not a good fit for the negative component of the 1-5 year-old group.

Figure 7.5 Fitted model varying the mean and standard deviation of the mixture model positive component by age group



7.8 Shape of underlying distributions

Both the negative and positive components of the mixture models shown so far were assumed to follow normal distributions. This may not hold so for the positive component, where there is antibody decay which may affect the fit in the key region where the two component densities meet. A more flexible model can be used instead, such as a *skew normal* distribution. The skew normal distribution is an extension of normal distribution allowing for a “skewness parameter”.

Assuming a normal distribution with constant mean and standard deviation for the negative component, and a skew normal distribution with varying parameters by age, Equation 7.12 becomes:

$$g_i(x) = (1 - \pi_i)f_N(x; \mu, \sigma) + \pi_i f_{SN}(x; v_i, \tau_i, \lambda_i),$$

Equation 7.16

for age groups $i = 1, \dots, \zeta$, f_{SN} being the skew normal density function

$$f_{SN}(x; v, \tau, \lambda) = \frac{2}{\tau} f_N(x; v, \tau) F_N\left(\lambda \frac{x-v}{\tau}\right),$$

Equation 7.17

where v , τ and λ are the location, dispersion and skewness parameters, respectively. The distribution is right skewed for $\lambda > 0$, it is left skewed for $\lambda < 0$ and it is normal for $\lambda = 0$. The mean μ of the skew normal distribution can be expressed as

$$\mu = v + \tau \frac{\lambda}{\sqrt{1 + \lambda^2}} \sqrt{\frac{2}{\pi}},$$

Equation 7.18

and the variance σ^2 as

$$\sigma^2 = \tau^2 \left(1 - \frac{2 \left(\frac{\lambda}{\sqrt{1 + \lambda^2}} \right)^2}{\pi} \right)$$

Equation 7.19

(Azzalini A, 1985; Roncalli T & Lagache T, 2004). Note that in Equations 7.18 and 7.19, π denotes the mathematical constant.

In order to take into account the censored observations as shown in the log-likelihood estimation from Equation 7.5 and Equation 7.6, the skew normal cumulative distribution function was calculated as follows:

$$F_{SN}(x) = 2 \int_{-\infty}^0 \int_{-\infty}^{\frac{x-u_1}{\tau}} F_2(u_1, u_2) du_1 du_2 ,$$

Equation 7.20

where $F_2(u_1, u_2)$ is a bivariate normal distribution with correlation $\rho = -\frac{\lambda}{\sqrt{1+\lambda^2}}$ (Azzalini A, 1985; Roncalli T & Lagache T, 2004; Wijsman RA, 1996).

For the VZV example, for $i = 1, \dots, 4$ age groups, the 18-parameter mixture model allowing for a skew normal positive component is shown in Figure 7.6 (negative component:

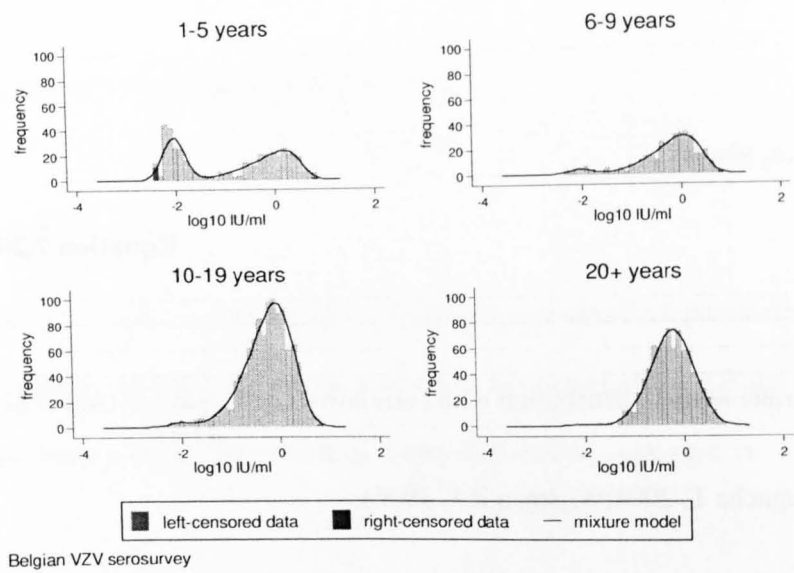
$\hat{\mu} = -2.09$, $\hat{\sigma} = 0.22$; positive component: $\hat{v}_1 = 0.49$, $\hat{v}_2 = 0.36$, $\hat{v}_3 = 0.24$, $\hat{v}_4 = 0.00$, $\hat{\tau}_1 = 0.83$,

$\hat{\tau}_2 = 0.76$, $\hat{\tau}_3 = 0.71$, $\hat{\tau}_4 = 0.54$, $\hat{\lambda}_1 = -4.12$, $\hat{\lambda}_2 = -3.2$, $\hat{\lambda}_3 = -2.61$, $\hat{\lambda}_4 = -1.07$; prevalence

estimates: $\hat{\pi}_1 = 0.59$, $\hat{\pi}_2 = 0.93$, $\hat{\pi}_3 = 0.98$, $\hat{\pi}_4 = 0.99$). Compared to the model shown in

Figure 7.5, the underlying skew normal distribution improves the fit in the higher measurements (positive component) for age groups 1-5, 6-9 and 10-19 years.

Figure 7.6 Fitted model with a skew normal positive component varying its parameters by age group



Using Equation 7.9, the likelihood ratio test between the models before and after including skewness parameters is defined as:

$$D_{df_2 - df_1} = -2(\ln(L_2) - \ln(L_1)) = -2(-2138.4 + 2091.6) = 93.6,$$

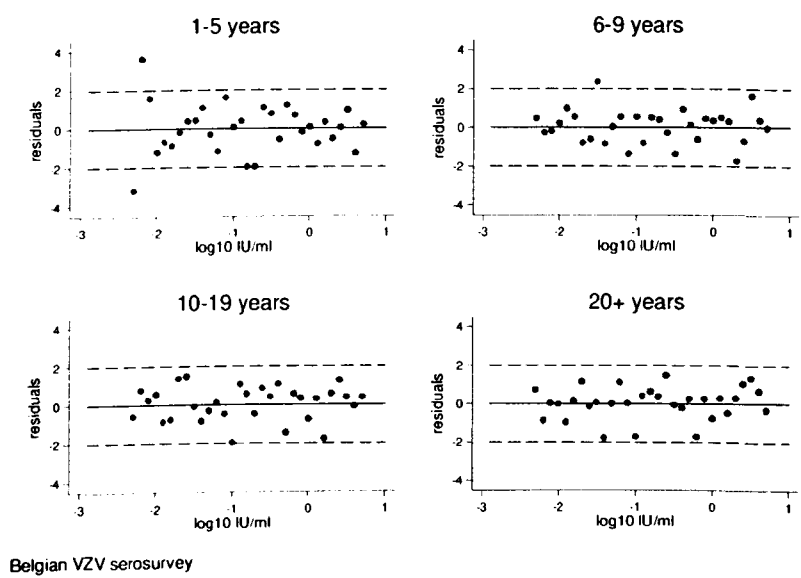
where L_2 is the model likelihood without the skewness parameter (Equation 7.15), and L_1 is the model likelihood including the skewness parameters (Equation 7.16). Under the null hypothesis that $\lambda = 0$, the statistic D follows an approximate chi-square distribution with $df_2 - df_1 = 36 - 32 = 4$ degrees of freedom, where df_2 and df_1 are the degrees of freedom for the model, excluding and including the skewness parameters, respectively. The highly significant p-value ($p < 0.001$) suggests that the model that includes the skewness parameters improves the model fit.

7.9 Model diagnostics (final model)

After grouping the data into bins in the same way as shown in Table 7.1, the Pearson chi-square test showed that there were borderline no significant differences between the data and the fitted distribution ($X^2 = 232.1$; $df = 201$; $p = 0.065$). After re-grouping the data, concentrating around the area of interest (-2 to -0.8), there was borderline significant evidence to reject the null hypothesis that the mixture model is correct ($X^2 = 20.2$; $df = 11$; $p = 0.043$).

Figure 7.7 shows there are no Pearson residuals outside the ± 2 limits for high antibody measurements. Therefore, the skew normal distribution for the positive component fits the data better, as compared to the previous example shown in Figure 7.2(b) where there were a number of residuals with extreme values particularly in age groups 1-5 and 6-9 years.

Figure 7.7 Pearson residuals by age group for the final model



Although the Kolmogorov-Smirnov test showed smaller differences between the observed and fitted data compared to the naïve model shown above, these were still highly significant ($D_{0.05,2746} = 0.10$; $p < 0.001$).

A comparison of the different mixture models is undertaken using AIC and BIC. A summary of the different models is given in Table 7.3. The mixture model using skew normal parameters fits the data better than the other two models. The additional number of parameters was taken into account since it had the lowest AIC and BIC values.

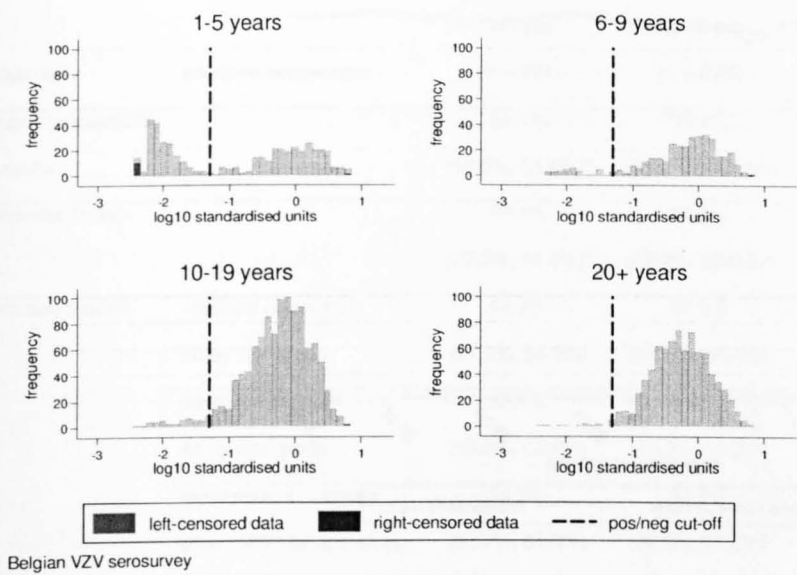
Table 7.3 AIC and BIC for the mixture models fitted

models	number of parameters	AIC	BIC
for both components -			
constant mean and sd by age group	8	4341.9	4389.3
for positive component -			
varying mean and sd by age group	14	4304.9	4387.8
for skew normal positive component -			
varying all parameters by age group	18	4219.3	4325.9

7.10 Seroprevalence estimation using standardisation

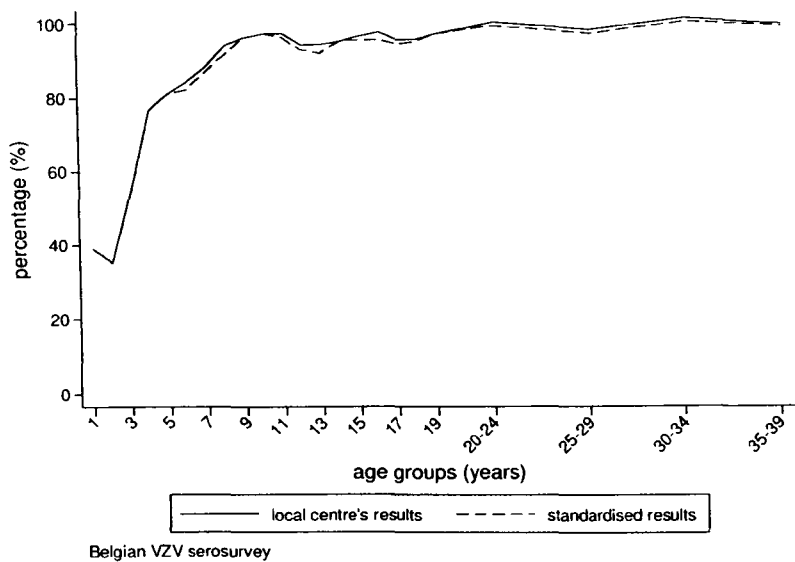
In order to demonstrate the difference between standardisation and mixture model methods, seroprevalence was also estimated after the serological results were standardised. The Belgium national serosurvey data were tested using the Enzygnost VZV (Dade Behring, Germany) assay, an assay also used by the Spanish reference centre (de Ory F *et al.*, 2006). The frequency distribution of the log₁₀-transformed results, together with the assay cut-off point (i.e. 0.05 IU/ml or -1.3 log₁₀ IU/ml), are shown in Figure 7.8.

Figure 7.8 Assay cut-off point and frequency distribution of VZV Belgium results by age group



Once the standardisation equation was applied to the serological results, these were classified into “positive” and “negative” according to the cut-off point of the reference centre’s assay (0.05 IU/ml). As shown in Figure 7.9 there were small differences between standardised and non-standardised seroprevalence estimates.

Figure 7.9 Comparison between standardised and non-standardised VZV seroprevalence estimates by age



7.11 Model comparisons

The seroprevalence estimates for each age group, using standardisation and the different mixture models, are presented in Table 7.4. The 95% confidence intervals around the age-specific seroprevalence estimates were estimated using the profile likelihood method. For each seroprevalence estimate, the maximum and minimum values for which the deviance was within 3.84 of the minimum, were calculated (Gay NJ, 1996; Vyse AJ *et al.*, 2006).

Table 7.4 Seroprevalence estimates by age groups for standardisation compared to different mixture models, with 95% CI

methods	positive component	age groups			
		1-9 yrs (n = 471)	10-19 yrs (n = 373)	20-29 yrs (n = 1109)	30-39 yrs (n = 809)
non-standardised		57.1%	90.4%	95.4%	98.25
results		(52.5%, 61.6%)*	(86.9%, 93.1%)*	(94.0%, 96.6%)*	(97.0%, 99.0%)*
standardisation		57.1%	89.0%	94.6%	97.3%
		(52.5%, 61.6%)*	(85.4%, 92.0%)*	(93.1%, 95.8%)*	(95.9%, 98.2%)*
mixture models	constant mean and	57.7%	92.0%	97.3%	99.2%
	sd by age group	(51.3%, 64.0%)	(87.4%, 95.4%)	(95.6%, 98.5%)	(97.9%, 99.8%)
	varying mean and	57.4%	91.8%	97.3%	99.2%
	sd by age group	(50.9%, 63.6%)	(87.2%, 95.3%)	(95.5%, 98.5%)	(97.8%, 99.8%)
	skew normal - varying	59.2%	93.4%	98.3%	99.3%
	parameters by age group	(52.7%, 65.5%)	(88.9%, 96.6%)	(96.9%, 99.3%)	(98.1%, 99.9%)

* 95% Binomial exact CI

In general, differences in seroprevalence between the methods were small. The standardisation method seroprevalence estimates were lower, whereas the skew normal model estimates were slightly higher when compared to the other methods, especially for ages 1-9 and 10-19 years.

7.12 Functional relationship of mean and standard deviation with age

In accordance with ESEN2 project guidelines, the Belgium laboratory tested approximately 100 samples for each 1-year age band less than 20 years of age. In addition to this, they tested approximately 200 samples for age groups 20-24, 25-29, 30-34 and 35-39 (Nardone A *et al.*, 2007). The aim was to obtain accurate seroprevalence estimates for narrow age groups, especially at younger ages.

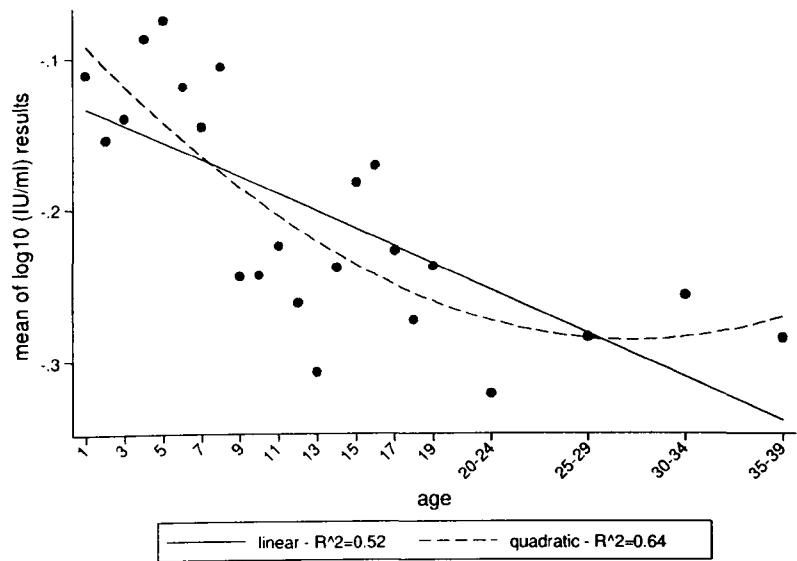
In order to obtain seroprevalence estimates for the 23 age groups, a mixture model of the type shown in Equation 7.15 would need 2 parameters for the negative component, 46 for

the positive (23 for the mean and 23 for the standard deviation) and 23 for the seroprevalence estimates resulting in 71 parameters. Similarly a 94-parameter model would be needed for a skew normal positive component (Equation 7.16). However, such a model could not be fitted in Stata, since a high number of parameters caused convergence problems for the maximum likelihood algorithm (Gould W & Sribney W, 1999).

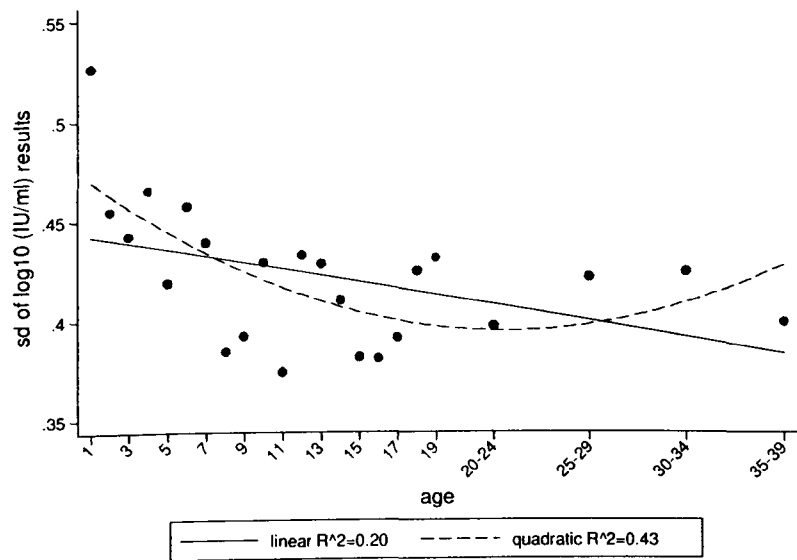
Separate seroprevalence estimates may be achieved for each of the age groups by reducing the number of parameters for the positive component of the distribution. To obtain this, the mean and the standard deviation were assumed to follow functional relationships by age. This is something that has been done in the past (Hardelid P, 2008), however, the difference in this case was that the age-specific seroprevalence estimates were not assumed to follow a linear relationship, since they were the subject of interest. To determine these functional relationships, all samples classified as “positive” by standardisation were initially selected. The mean and standard deviation of these measurements were plotted by age (Figure 7.10(a) and 7.10(b)). Both graphs suggest that a quadratic relationship may be more appropriate than a linear one to describe waning immunity as they provide a better fit.

Figure 7.10 Distribution of mean and standard deviation of \log_{10} IU/ml results by age confined to samples classified as “positive” by standardisation

(a) mean



(b) standard deviation



The model in Equation 7.15 then becomes

$$g_i(x) = (1 - \pi_i) f_N(x; \mu_-, \sigma_-) + \pi_i f_N(x; \mu_{+eq}(i), \sigma_{+eq}(i)),$$

Equation 7.21

where $\mu_{+eq}(i) = \mu_0 + i\mu_1 + i^2\mu_2$ and $\sigma_{+eq}(i) = \sigma_0 + i\sigma_1 + i^2\sigma_2$ for $i = 1, \dots, \zeta$ age groups.

Despite the reduction in the number of parameters, it was still not possible to obtain convergence of the ML algorithm after age was divided into 23 groups. The reason for this was that for the older age groups there were very small numbers of negative samples, and therefore, the negative component of the distribution could not be easily estimated. Table 7.5 shows the number of samples classified as “negative” by the standardisation method; for some age groups these were as low as 3 or 4.

Table 7.5 Frequency and percentage of negative samples per age stratum (according to how these were classified following standardisation)

age group	negative	%	Total
1	58	61.1%	95
2	61	64.9%	94
3	43	45.7%	94
4	22	23.4%	94
5	18	19.1%	94
6	17	18.1%	94
7	12	13.0%	92
8	8	8.5%	94
9	4	4.3%	93
10	3	3.3%	92
11	4	4.3%	93
12	7	7.4%	94
13	8	8.5%	94
14	5	5.4%	93
15	5	5.3%	94
16	5	5.3%	94
17	6	6.4%	94
18	10	5.7%	174
19	7	3.7%	187
20-24	4	2.0%	204
25-29	9	4.5%	200
30-34	3	1.5%	201
35-39	6	2.9%	204
Total	325	11.8%	2762

To achieve convergence, some of the older age groups in Table 7.5 were grouped into the following 14 age groups: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10-12, 13-14, 15-19, 20-29 and 30-39.

7.13 Functional relationship of skewness parameter with age

As shown above, a skew normal was considered a more flexible distribution to describe the shape of the positive component. To reduce the number of skewness parameters, a quadratic function for skewness was considered in addition to the functions used for the

location and dispersion parameters. This would transform Equation 7.16 into the following:

$$g_i(x) = (1 - \pi_i) f_N(x; \mu_-, \sigma_-) + \pi_i f_{SN}(x; \nu_{+eq}(i), \tau_{+eq}(i), \lambda_{+eq}(i)),$$

Equation 7.22

where $\nu_{+eq}(i) = \nu_0 + i\nu_1 + i^2\nu_2$, $\tau_{+eq}(i) = \tau_0 + i\tau_1 + i^2\tau_2$ and $\lambda_{+eq}(i) = \lambda_0 + i\lambda_1 + i^2\lambda_2$ for age groups $i = 1, \dots, \zeta$.

Using the 14 age groups shown in Table 7.5, the estimated parameters of the fitted model were:

Negative component: $\hat{\mu} = -2.11$, $\hat{\sigma} = 0.25$.

Positive component: $\hat{\nu}_0 = 0.56$, $\hat{\nu}_1 = -0.03$, $\hat{\nu}_2 = 0.00$, $\hat{\tau}_0 = 0.88$, $\hat{\tau}_1 = -0.01$, $\hat{\tau}_2 = 0.00$,
 $\hat{\lambda}_0 = -4.20$, $\hat{\lambda}_1 = 0.08$, $\hat{\lambda}_2 = 0.00$.

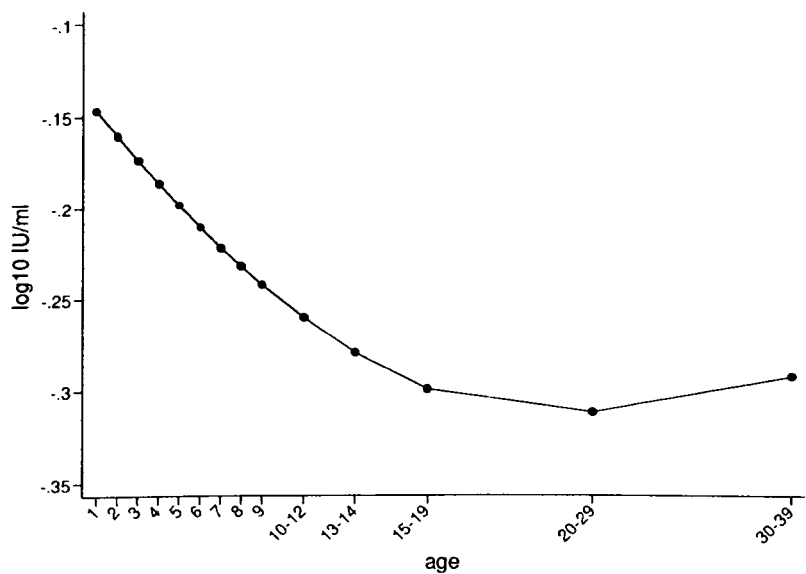
Prevalence estimates: $\hat{\pi}_1 = 0.40$, $\hat{\pi}_2 = 0.36$, $\hat{\pi}_3 = 0.56$, $\hat{\pi}_4 = 0.82$, $\hat{\pi}_5 = 0.84$, $\hat{\pi}_6 = 0.89$,
 $\hat{\pi}_7 = 0.91$, $\hat{\pi}_8 = 0.95$, $\hat{\pi}_9 = 0.98$, $\hat{\pi}_{10} = 0.99$, $\hat{\pi}_{11} = 0.97$, $\hat{\pi}_{12} = 0.98$, $\hat{\pi}_{13} = 0.99$ and $\hat{\pi}_{14} = 0.99$.

The model shown above was compared to a model with a normal underlying component for the positive group (instead of skew normal) using a likelihood ratio test. The skewness parameters were found to be significant ($D_4 = 99.8$; $p < 0.001$). Furthermore, the normal-components model had higher AIC and BIC ($AIC = 4243.5$ and $BIC = 4373.8$) compared to the skew normal component model ($AIC = 4149.7$ and $BIC = 4297.8$) which suggested that the latter provided a better fit.

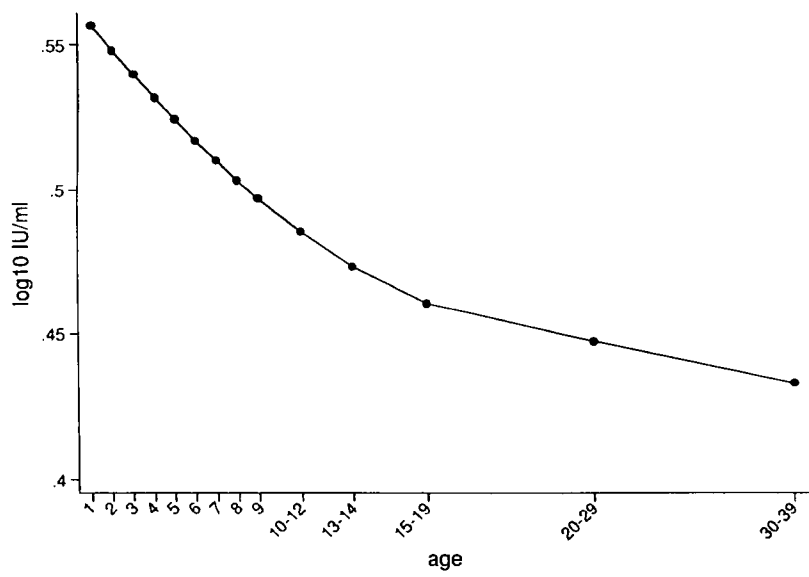
The mean and the standard deviation estimates by age group for the skew normal component as calculated by Equations 7.18 and 7.19 (using the location, dispersion and skewness parameters), respectively, are given in Figure 7.11.

Figure 7.11 Fitted mean and standard deviation of \log_{10} IU/ml results by age for the skew normal component of the mixture distribution (past infections)

(a) mean

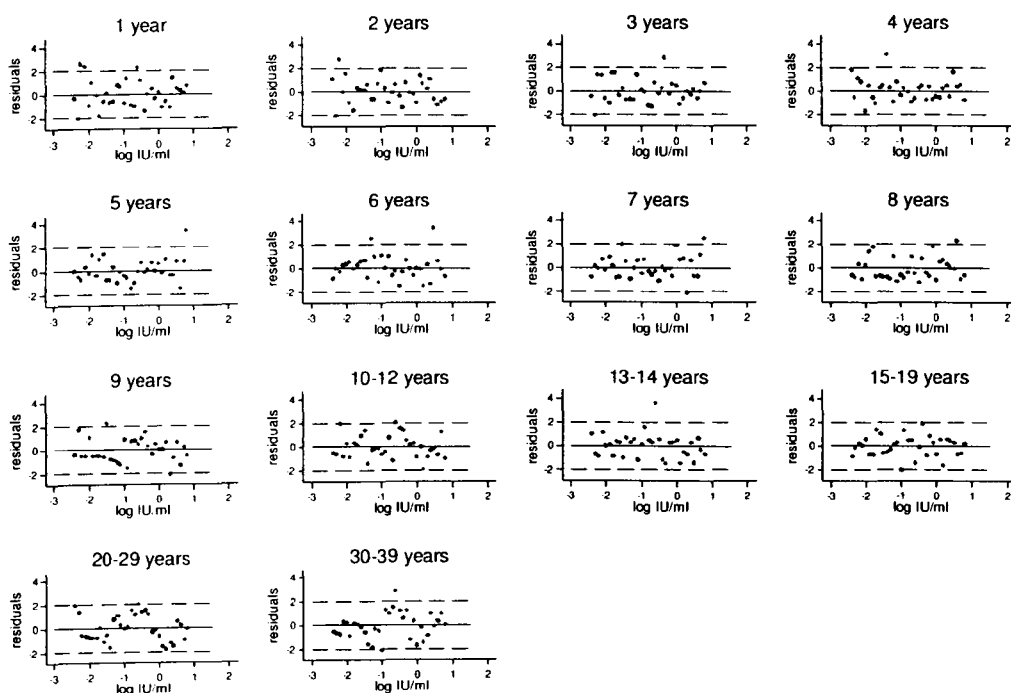


(b) standard deviation



The Pearson residuals are plotted for the 14 age groups in Figure 7.12 below. There are a few of the residuals outside the ± 2 range for age groups 1 to 9 years. These are within the range of -1.9 to -1 on the log₁₀-scale.

Figure 7.12 Pearson residuals vs. \log_{10} IU/ml by age group

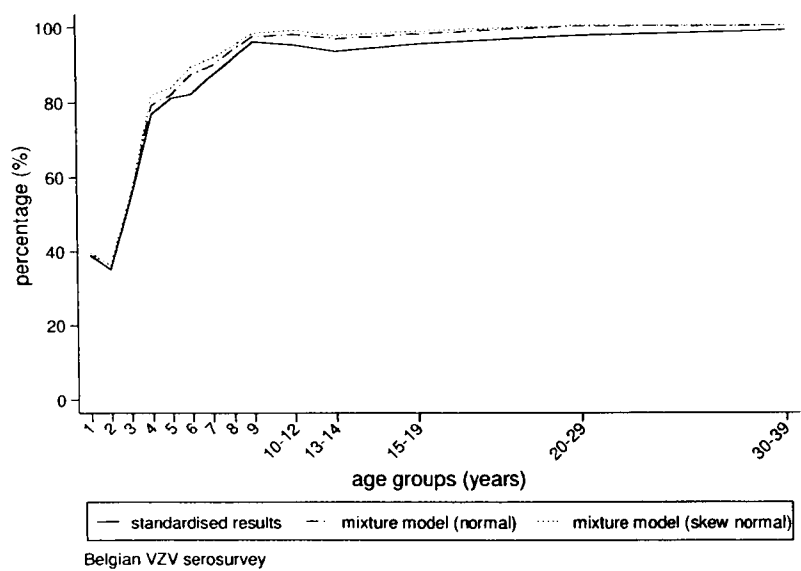


Belgian VZV serosurvey

7.14 Seroprevalence comparison for 14 age groups

The seroprevalence comparison between local results classification, classification after standardisation, and the two mixture models described above for 14 age groups are depicted in Figure 7.13. The estimated population seroprevalence using the skew normal model is consistently higher than the standardised seroprevalence for ages 4 and older. The highest difference between the two estimation methods occurs at the 6 years age group, where seroprevalence is estimated as 82% and 89% by the standardisation and the skew normal mixture model, respectively.

Figure 7.13 Seroprevalence comparisons



7.15 Conclusions

Besides using standardisation to estimate and compare population seroprevalence between samples tested in different national laboratories, an alternative approach is mixture models. Both these methods have a number of advantages and disadvantages.

As far as standardisation is concerned, the definition of the reference centre’s cut-off point has a big impact on the classification of the samples. A biased estimation of the cut-off point might lead to sample misclassification, and hence, biased seroprevalence estimates. The estimation of assay cut-offs points will be further discussed in Chapter 8.

On the other hand, although standardisation is a time-consuming and costly process, it provides a final standardised dataset of serological data. Apart from estimating seroprevalence, standardised results may also be applied to direct comparisons between quantitative serological data. In addition, more complex analyses may be undertaken on the

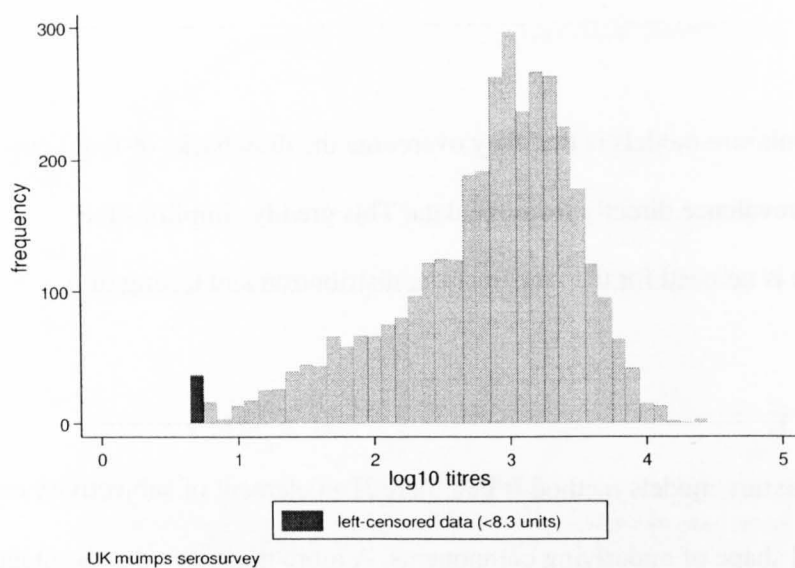
individual serum results, such as multivariable logistic regression and other types of modelling.

The main advantage of mixture models is that they overcome the drawbacks of fixed-cut-offs by estimating seroprevalence directly from the data. This greatly simplifies the methodology since there is no need for the construction, distribution and testing of a standardisation panel.

A disadvantage of the mixture models method is that there is an element of subjectivity on deciding the number and shape of underlying components. A more important disadvantage is the mixture models' dependency on the presence of separate peaks for the underlying populations. It is important to estimate the tails of the distributional components accurately around the area where the underlying distributions overlap, particularly in those situations where the components are not well-separated. Given that the shapes of underlying distributions are greatly affected by extreme measurements, the resulting estimates can be seriously biased. Hence, mixture model results should be interpreted with caution unless the underlying distributions are well separated.

An example illustrating this point is provided by a mumps serosurvey. A total of 3,575 samples were tested for mumps as part of the UK serosurvey carried out in 2000. The titre distribution of the standardised results, shown in Figure 7.14, has no separate peaks to indicate the underlying populations of those vaccinated or previously infected and those susceptible. Therefore, given that the underlying populations are not clearly defined, it is not possible to ensure that the positive and negative components are fitted correctly.

Figure 7.14 Distribution of the standardised mumps serosurvey results for UK



During this Chapter a simple example was chosen to demonstrate the application of a mixture model. However, as mentioned above, the construction of a mixture model is a subjective process, with respect to shape and number of underlying component distributions, which may potentially lead to model selection bias. Efforts have been undertaken in the past to introduce rules aiming to limit the level of subjectivity (Vyse AJ *et al.*, 2006;Biernacki C *et al.*, 2000;Hardelid P, 2008). However, despite these efforts, the element of subjectivity remains. The following quote is taken from Cormack: “Data-based model-selection is difficult and dangerous. The statistical question is usually phrased in terms of the significance levels at which a new parameter should be included in the model, or an existing parameter excluded. Measures such as Akaike’s criterion are advocated as objective decision rules. I suggest that biological understanding of the data set also plays a part...” (Morgan BJT, 2000).

A “biological understanding of the data” might include an appreciation, that for some antigens, measurement error might be an issue; for others, the presence of low positives (or

unprotected positives) owing to imperfect vaccines (as may be the case for mumps); waning antibodies with age; or lack of serological correlates of protection (as with pertussis), might all produce serological data for which it is not possible to separate negative and positive components. For such data, mixture modelling might lead to biased estimates, and so results need to be interpreted with caution.

Chapter 8: Estimating serological cut-offs for sero- epidemiology

8.1 Introduction

The material discussed in Chapters 4, 5 and 6 aimed at the validation and further development of the standardisation methodology used in the ESEN2 project, in order to achieve common units, and hence, comparable seroprevalence estimates. Once standardisation had been achieved, the results were classified into positive or negative, according to a cut-off point.

The estimation of an assay cut-off is very important, since it influences the classification of samples, and therefore, the national sero-profiles. The aim of this Chapter is to investigate the validity and impact of these cut-offs on the national sero-profiles, and to examine whether seroprevalence estimation can be improved by calculating alternative cut-offs.

8.2 Assay cut-offs as specified by the manufacturers

The cut-offs applied to the ESEN2 quantitative serological data were specified by the manufacturer of each reference country's assay.

Before assessing the validity of negative / positive cut-offs it is essential to understand how these were determined by assay manufacturers. Traditionally, fixed cut-offs are estimated from a sample of sera assumed to be negative i.e. from individuals who had neither been infected nor vaccinated in the past. A commonly used cut-off has been the sample mean of the log-transformed titres of a group of known negative samples plus 2 or 3 standard deviations. This ensures high specificity, disregarding however, the effect this may have on sensitivity (for a definition of sensitivity and specificity see Section 2.8) (Baughman AL *et al.*, 2006;Maple PAC *et al.*, 2006;Tong DDM *et al.*, 2007;Vyse AJ *et al.*, 1999;Parker RA *et al.*, 1990).

As was briefly mentioned in Chapter 7, mixture modelling has been used in the past on serological data, not for obtaining seroprevalence but for estimating positive / negative cut-offs (Parker RA *et al.*, 1990). Since then mixture models have commonly been used for estimating assay cut-offs (Maple PAC *et al.*, 2006; Sheppard C *et al.*, 2001). More details of this method will be given later in this Chapter.

Kernel density estimation has also been proposed in the past for estimating assay cut-offs.

This method involves obtaining a kernel density estimate of the distinction of both negative and positive samples, and setting the cut-off to equal one of the local minima.

However, this does not produce an objective and robust algorithm for determining a cut-off (Tong DDM *et al.*, 2007).

8.3 ROC curve analysis

Another method commonly used when comparing a new assay to an existing “gold standard” assay is the *receiver operating characteristic (ROC)* curve analysis. The sensitivity and specificity of the new assay can be calculated for every possible cut-off when compared to the established gold standard assay. Sensitivity results are then plotted against 1 - specificity and the *area under the curve (AUC)* provides a measurement of discrimination between the assays. The point that maximizes the sum of sensitivity and specificity may be chosen as a cut-off (i.e. $\max(\text{sensitivity} + \text{specificity})$) (Talukder Y *et al.*, 2005; Parker RA *et al.*, 1990; Hardelid P, 2008).

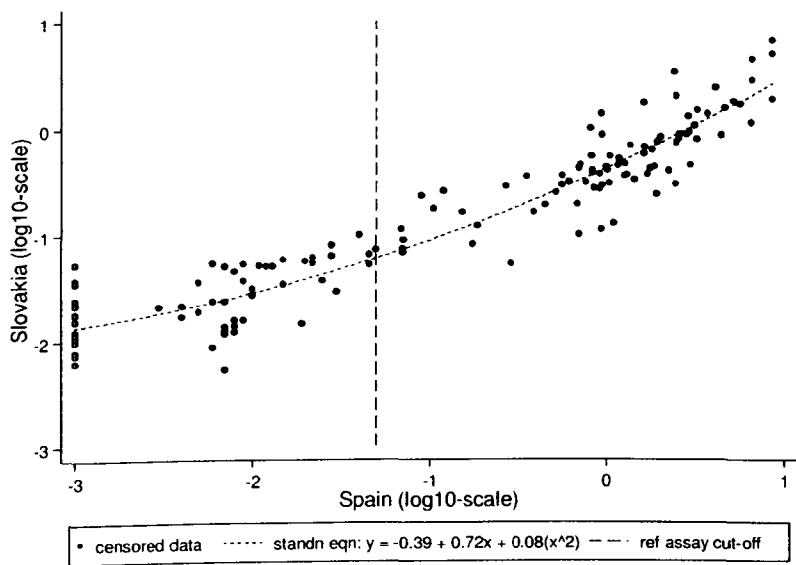
For the ESEN2 project, ROC curve analysis provides an alternative method for obtaining comparable outcomes without using regression equations to standardise serosurvey results.

ROC curves may be applied to the standardisation panel samples, by calculating sensitivity and specificity through the use of the reference centre’s qualitative results (defined

according to the reference assay's cut-off) for each testing laboratory's quantitative measurement. The value that maximises the sum of sensitivity and specificity may be chosen as the cut-off, to classify the testing laboratory's serosurvey results into positive and negative.

As an example, a panel of samples tested for VZV was prepared by the Spanish reference centre. The panel was subsequently sent to and tested by the Slovakian laboratory. The 145 samples tested by both laboratories are shown in Figure 8.1, together with the regression curve used to standardise the serosurvey results. The reference cut-off at 0.05 IU/ml (or -1.3 on the log₁₀-scale) classified the standardised units into positive and negative.

Figure 8.1 Standardisation curve for data from Slovakia and the Spanish reference centre



For the ROC curve analysis, the sensitivity and specificity were calculated for cut-offs placed between successive Slovakian panel measurements. It is clear from Table 8.1 that for -1.16 log₁₀-transformed units (0.07 IU/ml) the sum of sensitivity and specificity is

maximised. Using such a cut-off point for the Slovakian assay the sensitivity is 98% and the specificity 97%.

Table 8.1 Sensitivity and specificity for different cut-off points for Slovakia

cut-off for Slovakia	sensitivity	specificity	sensitivity + specificity
-1.46	100.0	63.3	163.3
-1.40	100.0	71.7	171.7
-1.34	100.0	73.3	173.3
-1.28	100.0	81.7	181.7
-1.22	98.8	93.3	192.2
-1.16	97.6	96.7	194.3
-1.09	94.1	98.3	192.5
-1.03	92.9	98.3	191.3
-0.97	91.8	100.0	191.8

* only some of the potential cut-offs are given in this Table

The validity of the cut-off as estimated using the ROC curve depends on how well the positives are separated from the negatives. The separation between negatives and positives for the standardisation panel can be assessed by plotting sensitivity against 1 – specificity and calculating the AUC. Given that *AUC* takes values between 0.5 and 1, an *AUC* less than 0.75 may be interpreted as poor discrimination between the two assays.

Figure 8.2 shows the sensitivity plotted against 1 - specificity for the standardisation panel shown above. Here *AUC* = 0.996 which means that there is very good separation between positive and negative results.

Figure 8.2 Area under the curve for Slovakia VZV panel

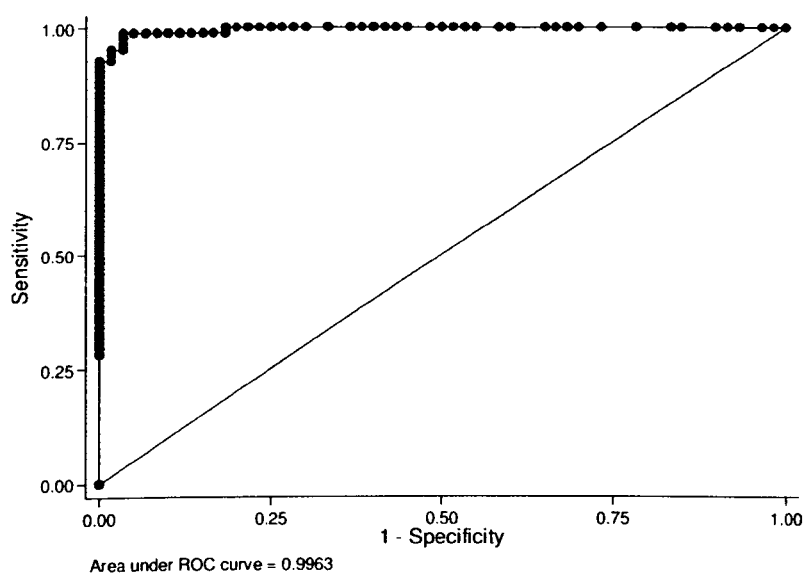
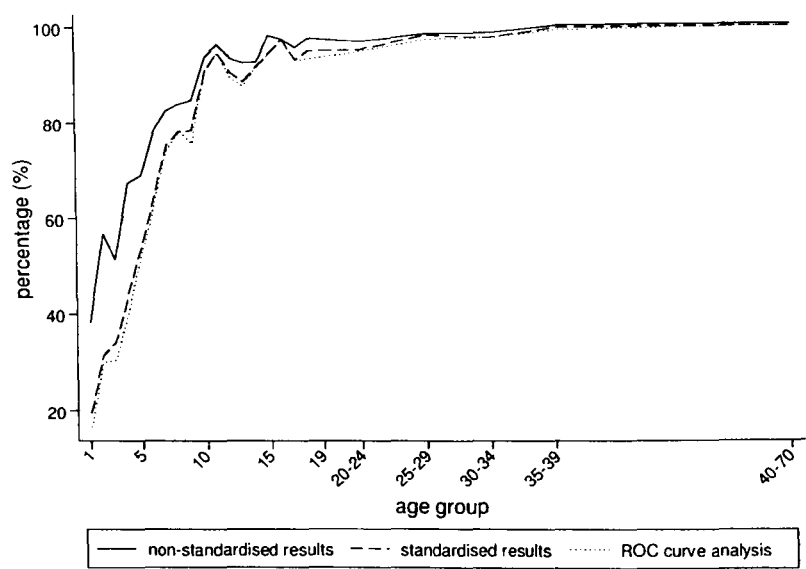


Figure 8.3 shows a comparison between the estimated seroprevalence for Slovakia (a) using the non-standardised serosurvey and the local assay cut-off, (b) the standardised serosurvey results and the reference centre’s cut-off as used in the ESEN2 project and (c) the non-standardised serosurvey results and the estimated cut-off using ROC curve analysis. The seroprevalence estimates based on non-standardised results are overestimated, especially for ages less than 10 years. The method based on ROC curves gave very similar estimates to the standardisation method used for the ESEN2 project.

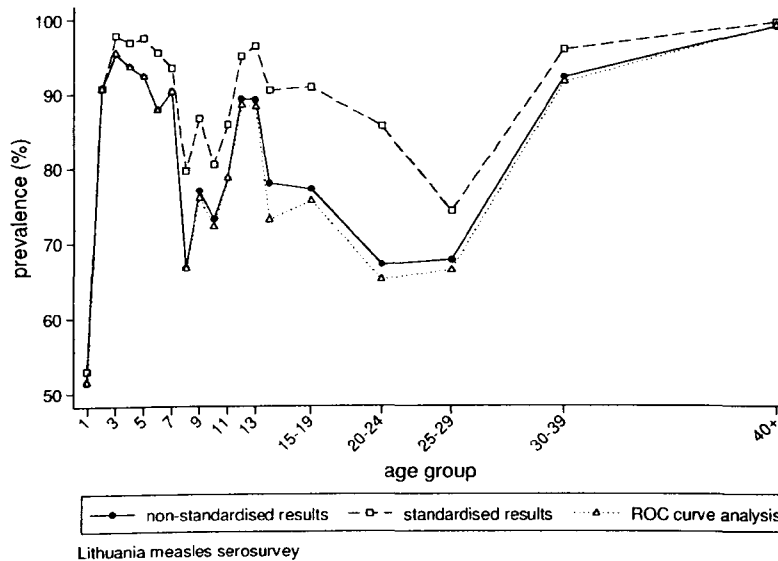
Figure 8.3 Seroprevalence estimates for Slovakia using local assay results, standardisation and ROC curve analysis



However, it needs to be noted that the AUC is a measurement of discrimination for the standardisation panel and not the serosurvey results which is where the estimated cut-off is being applied to. In general, for the standardisation panels, there were small number of measurements around the area of the negative / positive cut-off which meant there was little chance for misclassification and therefore, high AUC estimates. However, there were often high numbers of data around the critical area around the cut-off when looking the national serosurvey.

To demonstrate this point, the Lithuania measles example introduced in Figures 4.9 and 4.10 is being revisited. Following a ROC analysis, the cut-off was estimated as -0.80 (on the log₁₀-scale) with very high separation between positives and negatives ($AUC = 0.995$). However, looking at Figure 4.9, it is clear that such a cut-off is overestimated for the Lithuanian assay. As shown in Figure 8.4, the ROC method does not adjust well for the different assay scale resulting in seroprevalence close to the non-standardised estimates.

Figure 8.4 Seroprevalence estimates using ROC analysis for Lithuanian measles results



The main advantage of the ROC analysis described here is that it is a non-parametric method and therefore, does not depend on regression assumptions. Another benefit is that it provides an easier method to implement, avoiding subjective decisions such as the type of equation used.

The main drawback of the ROC analysis is that it does not make full use of the dataset, depending very much on how clear the distinction is between positive and negative results around the area of interest. In scenarios where no clear distinction exists between positive and negative sample for the national serosurveys as was often the case for the ESEN2 project, this method is not expected to work.

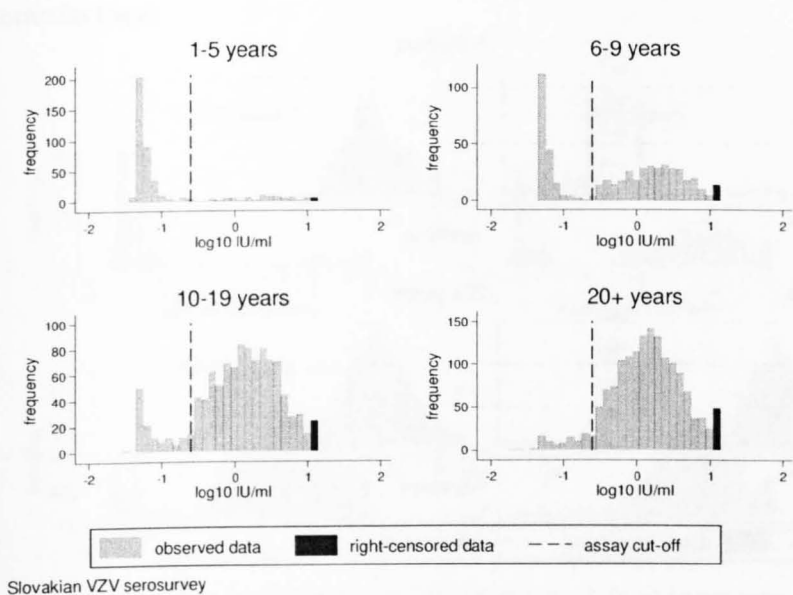
8.4 Are assay cut-offs appropriate for sero-epidemiology?

Serological assays are designed for clinical purposes and therefore, assay cut-offs are estimated to suit diagnosis in individuals. In terms of vaccine-preventable infections, the aim of the assay manufacturer is usually to ensure that all truly negative samples are classified as “negative” by the test. In other words, a cut-off point aims at maximising specificity, disregarding the effect this might have on sensitivity. This may lead to some protected individuals being classified as “susceptible” (Ades AE, 1990; Vyse AJ *et al.*, 2006; Hardelid P, 2008).

It has been claimed in the past that highly specific tests are important in epidemiology, since false-positive results might cause “false” outbreak investigations or unnecessary interventions such as antibiotic treatments (Baughman AL *et al.*, 2006). However, while highly specific tests may be desirable in some cases, the requirements for estimating seroprevalence in large population studies are different. In this case, the aim should be to minimise the bias in seroprevalence. Using a specificity-oriented assay cut-off may result in underestimating seroprevalence (Morgan BJT, 2000; Vyse AJ *et al.*, 2004; Hardelid P, 2008).

The Slovakian VZV serosurvey collected in 2002 provides an example of this. The Slovakian laboratory used the Euroimmune assay to test their samples with a negative / positive cut-off as defined by the assay manufacturer at 0.25 IU/ml or -0.6 on the \log_{10} -scale. By observing the distribution of the log-transformed measurements in Figure 8.5, the cut-off point seems to lie slightly too high on the left tail of what might be expected to be the distribution of positive samples.

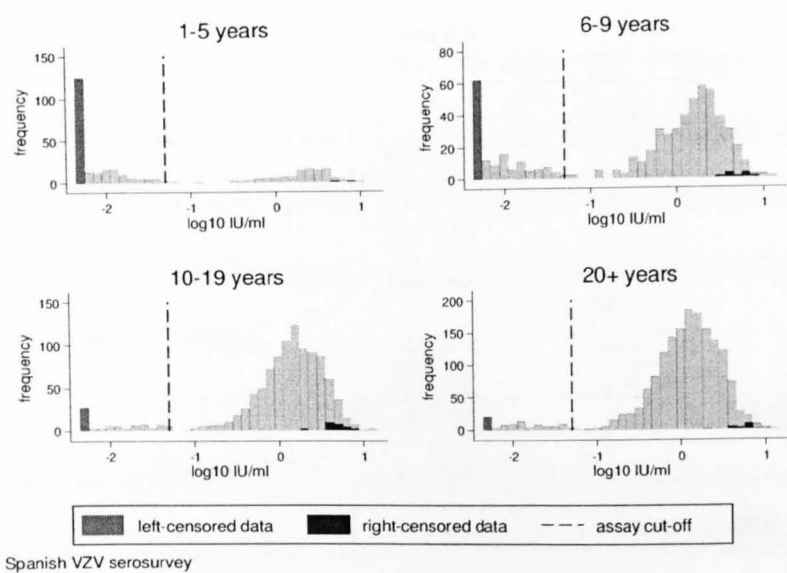
Figure 8.5 Distribution of non-standardised serological VZV results for Slovakia and positive / negative cut-off point as defined by the assay manufacturer



Note that a number of samples tested by the Slovakia laboratory fall outside the detection range of the assay. For some but not all of these samples quantitative results were obtained. The qualitative results are shown in the graph after the method of simple substitution had been used (Chapter 5), thus generating the spike in Figure 8.5.

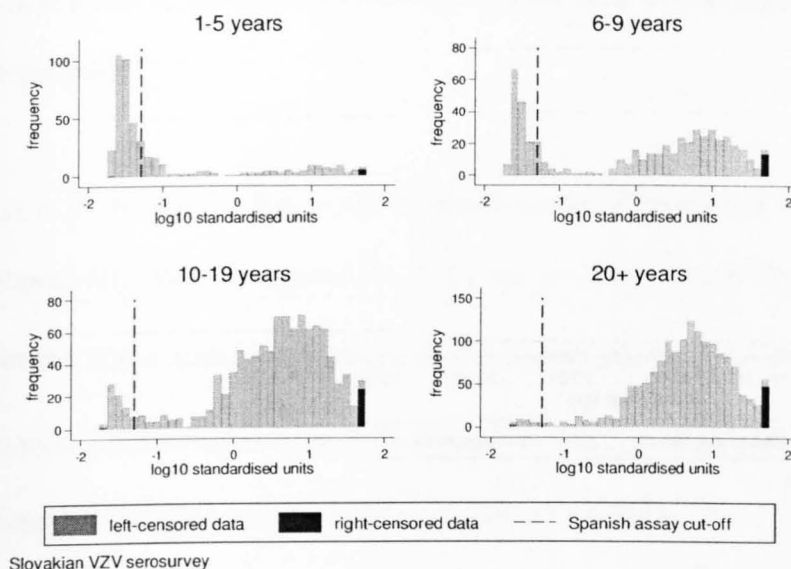
In the above example, the emphasis has been given to the assay specificity. However, this was not the case for the Dade Behring assay used by the reference centre to test Spain's national serosurvey in 2002. In Figure 8.6, the assay cut-off seems to overlap slightly with the distribution of the negative samples (cut-off = 0.05 IU/ml or -1.3 on the log₁₀-scale for the Dade Behring assay), especially in samples taken from children.

Figure 8.6 Distribution of serological VZV results for Spain and positive / negative cut-off point as defined by the assay manufacturer



For the purposes of the ESEN2 project it was essential that this cut-off was well-chosen, since the laboratory that implemented the testing (i.e. Spain) was the reference laboratory for the VZV work package. This meant that this particular cut-off was used to re-classify the standardised results from all countries into positive and negative. For example, once the Slovakian serosurvey results shown in Figure 8.5 had been standardised, the reference centre's cut-off was used. Figure 8.7 shows the standardised Slovakian results, together with the cut-offs as chosen by the Spanish assay manufacturer. The chosen cut-off seems clearly to overlap with the distribution of the negative results.

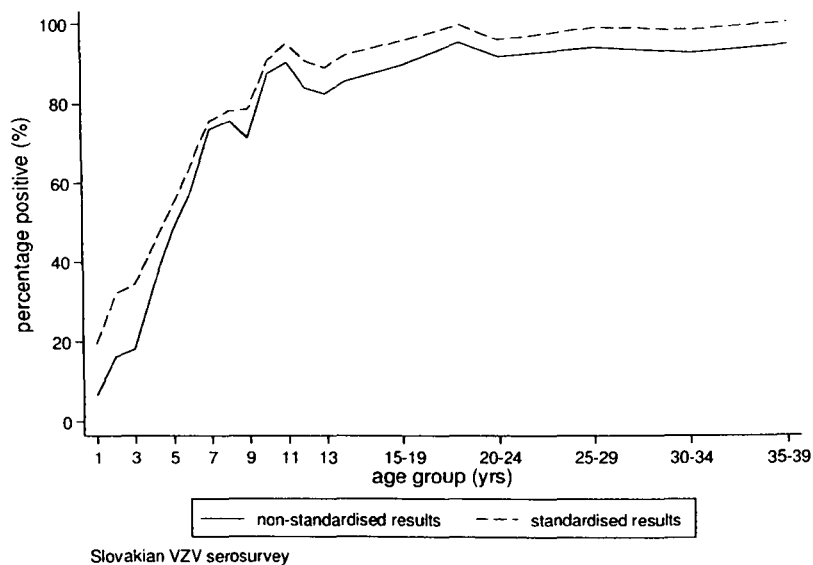
Figure 8.7 Distribution of standardised serological VZV results for Slovakia and positive / negative cut-off point as defined by the reference centre's assay manufacturer



Considering the standardised results, the cut-off stands in the right tail of what would be expected as the distribution of the negative results. This contrasts with the non-standardised results, where the local cut-off was on the left tail of what one would expect to be the distribution of positive results (Figure 8.5). Therefore, a large number of negative samples are expected to be re-classified as positive after the results have been standardised. Given that some of these results may be falsely re-classified as positives, the standardisation method may not have worked very well in this example.

The estimated seroprevalence by age using both non-standardised (cut-off = 0.25 IU/ml) and standardised measurements (cut-off = 0.05 IU/ml) are given in Figure 8.8.

Figure 8.8 Standardised vs non-standardised seroprevalence estimates for Slovakia
serosurvey



8.5 Estimating assay cut-offs using mixture models

An assay cut-off that minimises the misclassification between positive and negative results is most suitable for estimating population seroprevalence. One way to achieve such an objective cut-off point is through mixture modelling (Baughman AL *et al.*, 2006; Talukder Y *et al.*, 2005; Talukder Y *et al.*, 2007; Dong Z, 1997; Parker RA *et al.*, 1990).

Assuming a simple mixture model with two underlying populations of positive (past infections) and negative samples (uninfected individuals), a two-component mixture model can be fitted.

Similar to the mixture model shown in Equation 7.3, assuming for simplicity there is no age effect, let f_- and f_+ be the density functions of the negative and the positive component, respectively. Then a simple mixture model can be expressed in the following form:

$$g(x) = (1 - \pi)f_-(x; \theta_-) + \pi f_+(x; \theta_+),$$

Equation 8.1

where θ_- and θ_+ are the model parameters of the negative and positive components, respectively.

Let F_- , F_+ be the cumulative distribution functions of the density functions f_- and f_+ , respectively. Then for a given cut-off C_0 , the proportion of incorrectly classified positive samples (false negative rate) is $F_+(C_0; \hat{\theta}_+)$ and the proportion of incorrectly classified negatives (false positive rate) is $1 - F_-(C_0; \hat{\theta}_-)$, where $\hat{\theta}_-$ and $\hat{\theta}_+$ are the parameters estimated using the mixture model in Equation 8.1 above.

One approach is to estimate a cut-off C_0 that minimizes the total misclassification $h(C_0)$ defined as the sum of the false positive and false negative rates which is equivalent to:

$$h(C_0) = (1 - F_-(C_0; \hat{\theta}_-)) + F_+(C_0; \hat{\theta}_+).$$

Equation 8.2

A drawback of Equation 8.2 is that it does not take into account the proportion of samples attributed to each component. The total proportion of misclassified samples depends on the estimated seroprevalence $\hat{\pi}$ derived from Equation 8.1. Thus, to minimize the total number of misclassified samples, one can find the cut-off point that minimizes a weighted version of Equation 8.2, namely:

$$h(C_0) = (1 - \hat{\pi})(1 - F_-(C_0; \hat{\theta}_-)) + \hat{\pi}F_+(C_0; \hat{\theta}_+)$$

Equation 8.3

(Baughman AL *et al.*, 2006).

Other cut-off points that present more “conservative” or more “liberal” estimates than the cut-off described in Equation 8.3 have been proposed (Tong DDM *et al.*, 2007).

An alternative option is to seek a cut-off for which the seroprevalence bias is zero, i.e. returns equal numbers of false negative and false positive results. This occurs when

$$(1 - \hat{\pi}) \left(1 - F_{-}(C_0; \hat{\theta}_{-}) \right) = \hat{\pi} F_{+}(C_0; \hat{\theta}_{+}),$$

which can be achieved by solving:

$$\left| \hat{\pi} F_{+}(C_0; \hat{\theta}_{+}) - (1 - \hat{\pi}) \left(1 - F_{-}(C_0; \hat{\theta}_{-}) \right) \right| = 0.$$

Equation 8.4

The purpose of re-estimating the negative / positive cut-off of an assay, is to eliminate the bias arising from the assay manufacturer’s choice of cut-off that is designed for diagnostic purposes. In this way, the cut-off proposed in Equation 8.4 is a unique cut-off adjusted for the specific purposes of this project. Such unique fixed cut-offs, however they are estimated, suffer from a serious limitation: they cannot adjust for the important change in seroprevalence that naturally occurs with age (Vyse AJ *et al.*, 2004; Vyse AJ *et al.*, 2006). Therefore, an alternative method is proposed that results in different cut-off estimates by age group. Age-specific cut-offs \hat{C}_{0i} can be estimated after substituting the parameter estimates $\hat{\pi}$, $\hat{\theta}_{-}$ and $\hat{\theta}_{+}$ by $\hat{\pi}_i$, $\hat{\theta}_{-i}$ and $\hat{\theta}_{+i}$, for age groups $i = 1, \dots, \zeta$. Equation 8.4 then becomes:

$$\left| \hat{\pi}_i F_+(C_{0i}; \hat{\theta}_{+i}) - (1 - \hat{\pi}_i) (1 - F_-(C_{0i}; \hat{\theta}_{-i})) \right| = 0.$$

Equation 8.5

The weighted cut-off in Equation 8.5 is considered to produce the least biased cut-off for the objectives of the project and will mainly be used thereafter.

In order to find the root of the continuous monotonic functions expressed in Equations 8.4 and 8.5, the optimize command was used in R (Brent R, 1973). To avoid convergence problems the minimization procedure was restricted to a “sensible” area around, e.g. between the location parameters of the underlying component distributions.

Note that the cut-off estimation method shown here was applied to a mixture model with two underlying components but can easily be extended to three or more e.g. negative, vaccinated and previously infected individuals. For such a scenario, the two (negative / vaccinated and vaccinated / past infection) cut-offs can be estimated after adding a third component to Equations 8.4 and 8.5 and restricting appropriately the optimisation procedure.

8.6 Estimating an age-specific cut-off for the Spanish VZV serosurvey

As described above, the assay cut-off used by the VZV reference centre (Spain), seemed to overlap with the underlying distribution of the negative results (Figure 8.6). An alternative cut-off was estimated using mixture modelling.

Assuming the existence of two underlying populations - one for previously infected and one for negative individuals - a two-component mixture model was constructed. A normal

distribution was assumed for individuals with no past infections and a skew normal for those previously infected. Note that, the mixture model with skew normal component was compared to the model with a normal component, using a likelihood ratio which showed that the added skewness parameters significantly improved the model fit ($D_4 = 99.8$; $p < 0.001$).

Assuming waning immunity for the positive component, a mixture model of the type described in Equation 7.16 was used, where $i = 1, \dots, 4$ represents the age groups 1-4, 5-9, 10-19 and 20+ years, respectively. The left- and right-censored observations were taken into account as shown in Equations 7.5 and 7.6, respectively.

The resulting estimates were:

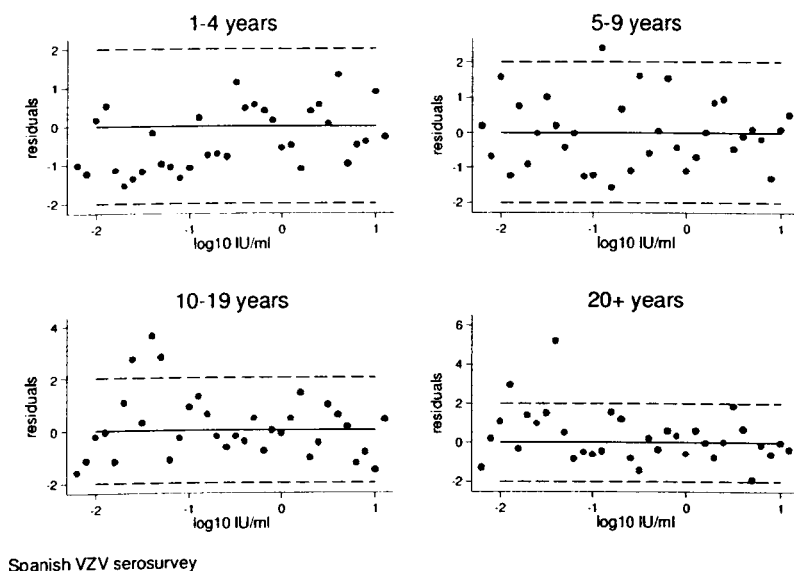
Negative component: $\hat{\mu}_- = -2.21$, $\hat{\sigma}_- = 0.50$.

Positive component: $\hat{v}_{+1} = 0.63$, $\hat{v}_{+2} = 0.52$, $\hat{v}_{+3} = 0.45$, $\hat{v}_{+4} = 0.38$, $\hat{\tau}_{+1} = 0.48$, $\hat{\tau}_{+2} = 0.52$, $\hat{\tau}_{+3} = 0.49$, $\hat{\tau}_{+4} = 0.49$, $\hat{\lambda}_{+1} = -2.81$, $\hat{\lambda}_{+2} = -2.19$, $\hat{\lambda}_{+3} = -1.69$, $\hat{\lambda}_{+4} = -1.62$.

Prevalence estimates: $\hat{\pi}_{+1} = 0.33$, $\hat{\pi}_{+2} = 0.75$, $\hat{\pi}_{+3} = 0.93$, $\hat{\pi}_{+4} = 0.94$.

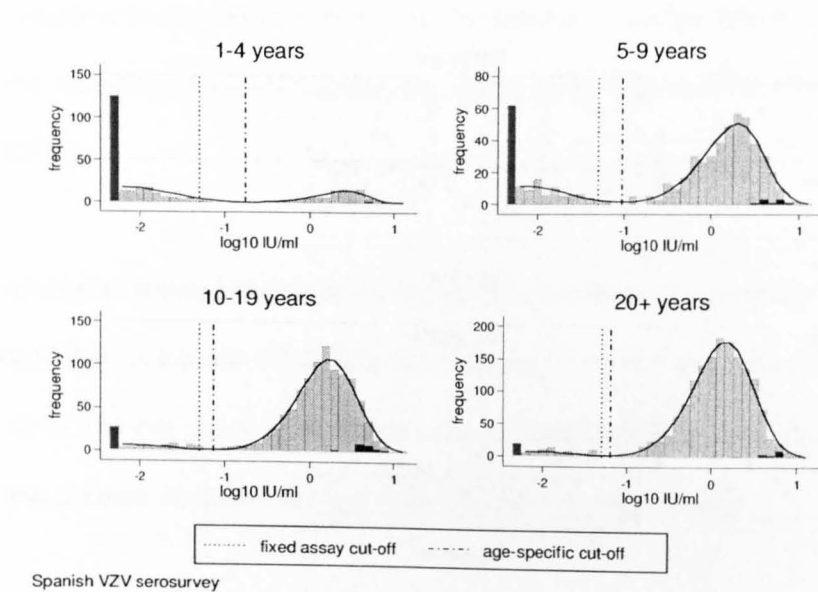
The residual plot in Figure 8.9 revealed a few residuals outside the ± 2 range for low serological measurements, for age groups 10-19 and more than 20 years.

Figure 8.9 Pearson residuals for Spanish VZV serosurvey plotted against \log_{10} IU/ml by age group



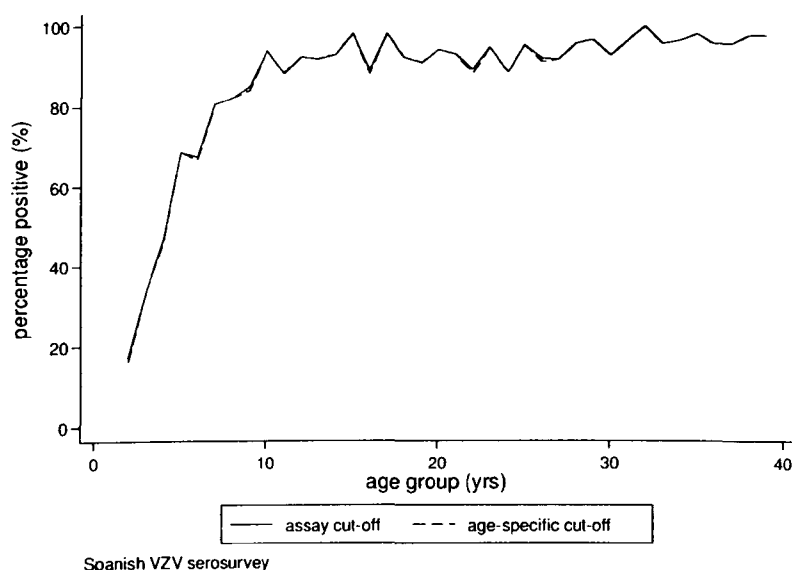
After applying Equation 8.5, the estimated cut-offs for the age groups were $\hat{C}_{01} = -0.76$, or 0.17 IU/ml for ages 1-4, $\hat{C}_{02} = -1.03$, or 0.09 IU/ml for ages 5-9, $\hat{C}_{03} = -1.13$, or 0.07 IU/ml for ages 10-19 and $\hat{C}_{04} = -1.20$, or 0.06 IU/ml for ages 20+ (The estimation program in R is given in Appendix I (D)). The observed data and the fitted distribution are given in Figure 8.10 below. The Spanish cut-off, as defined by the assay manufacturer (-1.3 on the \log_{10} -scale), and the age-specific cut-offs derived from the mixture models are also shown. The age-specific cut-off is much higher than the fixed cut-off for the younger age group (1-4 years). The age-specific cut-offs are closer to the fixed value for the older age groups.

Figure 8.10 Comparison between the fixed cut-off given by the assay manufacturer and the age-specific cut-off as estimated using the mixture model estimates



Despite the difference in cut-off estimates, seroprevalence estimates are almost identical for the two methods (Figure 8.11). This is a result of the clear separation between the underlying distributions, since in this case there are few observations that can be potentially misclassified.

Figure 8.11 Seroprevalence comparisons using different cut-off estimates



8.7 Application of the estimated cut-offs to other national serosurveys

As explained above, weighted cut-offs provide a better estimate than non-weighted ones, since they take into account the proportion of samples in each component distribution, and therefore, minimise misclassification. However, they cannot be applied directly to other standardised serosurveys, due to their dependency on the component distribution proportions at the reference centre.

The following method is proposed that makes it possible to apply a mixture model estimated cut-off to a serosurvey, without actually fitting a mixture model to these data (given that the data have initially been standardised):

1. A mixture model is fitted to the reference centre's data as shown in the example in the previous Section.

2. Rough estimates of the proportion of the testing laboratory serosurvey samples for each component are obtained using the assay cut-off or the reference centre's cut-off (since the data have been standardised).

3. The mixture model parameters from Step 1 are used in combination with the seroprevalence estimates obtained from Step 2 to estimate a cut-off for the national assay using Equation 8.5.

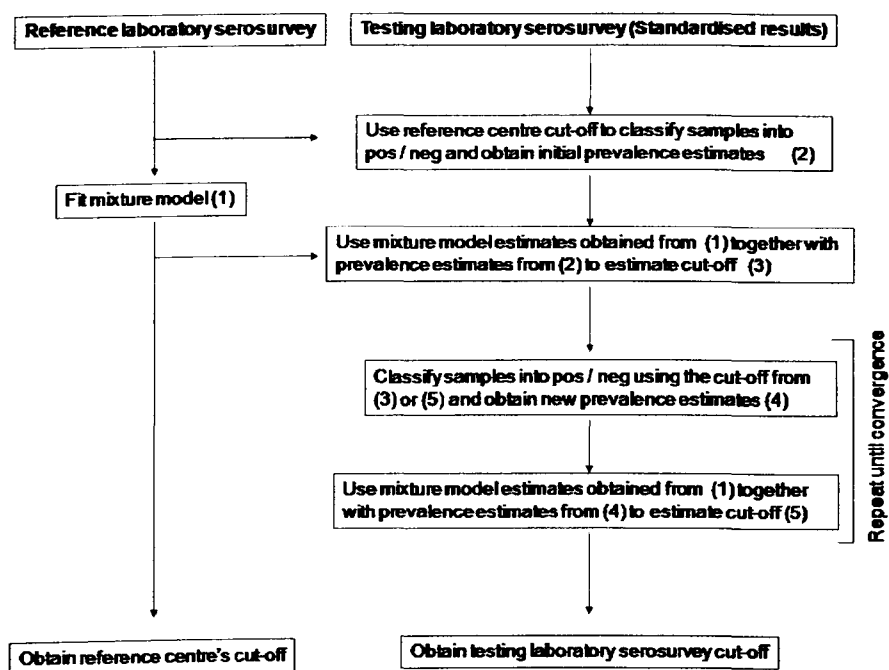
4. The cut-off is used to re-classify samples and obtain new seroprevalence estimates.

5. The mixture model parameters from Step 1 are used again in combination with the seroprevalence estimates obtained from Step 4 to obtain a new estimated cut-off.

Steps 4 and 5 may be repeated in an iterative process until converged to a final cut-off estimate.

The process for estimating a cut-off for both reference and national serosurveys is shown in Figure 8.12 below.

Figure 8.12 Flow-chart showing the process of estimating a cut-off point



There is a key assumption which must be satisfied if the iterative method is to succeed, namely that the component distributions should be the same at the reference centre and the testing laboratory serosurveys.

The Slovakian standardised serosurvey results presented previously will serve as an example of cut-off estimation. Using the Spanish reference centre's assay cut-off (0.05 IU/ml), seroprevalence estimates were obtained for Slovakia by age group ($\hat{\pi}_1 = 0.33$ for ages 1-4, $\hat{\pi}_2 = 0.70$ for 5-9, $\hat{\pi}_3 = 0.93$ for 10-19 and $\hat{\pi}_4 = 0.98$ for 20+). These seroprevalence estimates, in combination with the mixture model parameter estimates obtained previously (negative component: $\hat{\mu}_- = -2.21$, $\hat{\sigma}_- = 0.50$; positive component: $\hat{\nu}_{+1} = 0.63$, $\hat{\nu}_{+2} = 0.52$, $\hat{\nu}_{+3} = 0.45$, $\hat{\nu}_{+4} = 0.38$, $\hat{\tau}_{+1} = 0.48$, $\hat{\tau}_{+2} = 0.52$, $\hat{\tau}_{+3} = 0.49$, $\hat{\tau}_{+4} = 0.49$, $\hat{\lambda}_{+1} = -2.81$, $\hat{\lambda}_{+2} = -2.19$, $\hat{\lambda}_{+3} = -1.69$, $\hat{\lambda}_{+4} = -1.62$), allowed for age-specific cut-offs to be obtained (Equation 8.5) by the iterative process described above.

Table 8.2 presents a comparison of the different methods for estimating assay cut-offs for the Slovakian (standardised) assay.

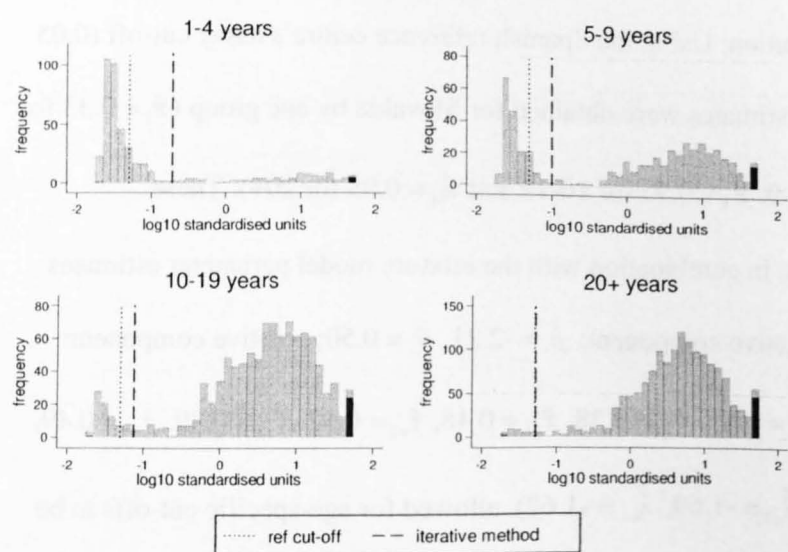
Table 8.2 Comparison between the cut-offs as estimated by the different methods

Methods	age groups			
	1-4 yrs	5-9 yrs	10-19 yrs	20+ yrs
Slovakian assay cut-off				
(applicable to non-standardised results)	0.25 (-0.60)	0.25 (-0.60)	0.25 (-0.60)	0.25 (-0.60)
Spanish ref assay cut-off				
(used in ESEN2 project)	0.05 (-1.30)	0.05 (-1.30)	0.05 (-1.30)	0.05 (-1.30)
mixture model cut-off (iterative method)				
(Equation 8.5)	0.19 (-0.72)	0.10 (-1.00)	0.08 (-1.12)	0.05 (-1.29)

* cut-offs in IU/ml. The log₁₀ transformation is given in brackets.

The standardised results in Figure 8.13 show that the reference centre cut-off is placed at a very low point, well within the tail of the underlying distribution of non-infected individuals. The iterative method gave higher estimates for ages less than 20 years.

Figure 8.13 Comparison between the cut-offs as estimated by the different methods

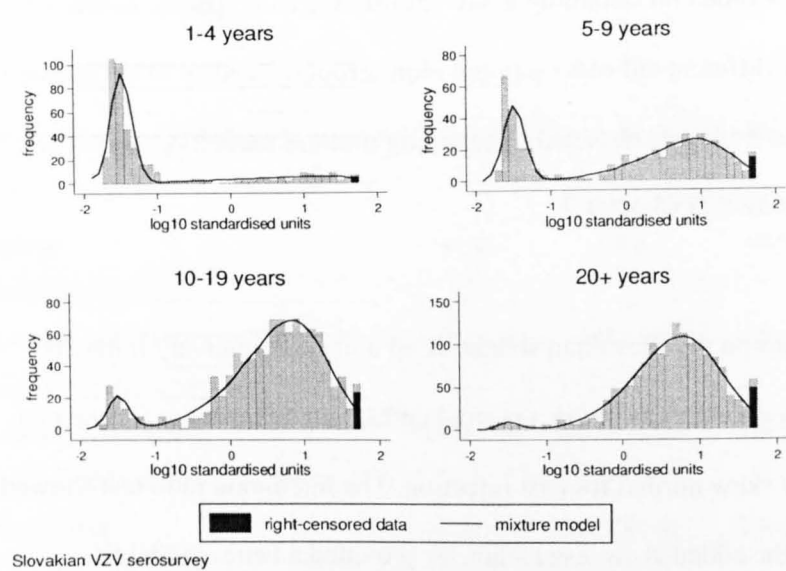


8.8 Impact of cut-off estimates on population seroprevalence

In Section 8.7, different methods for obtaining assay cut-offs were compared. In this Section, the impact of the different cut-offs on population seroprevalence will be assessed. These will be compared to estimates obtained by applying mixture modelling directly on the Slovakian results as shown in Chapter 7.

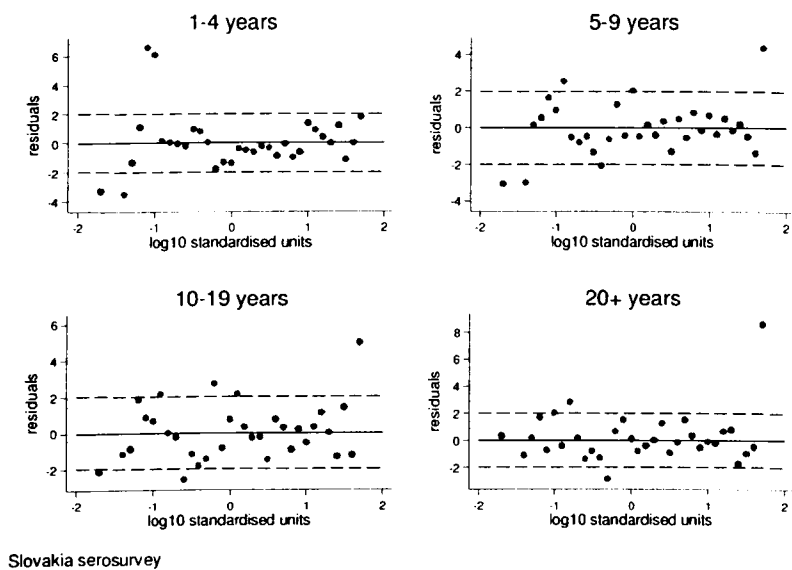
A mixture model was fitted on the Slovakian standardised units (alternatively it may be applied to non-standardised units), assuming a normal underlying component for the non-infected individuals and a skew normal for past infection. The likelihood ratio test showed significant evidence that the added skewness parameter provided a better model fit ($D_4 = 83.9$; $p < 0.001$). The estimated model is shown in Figure 8.14 (negative component: $\hat{\mu}_- = -1.54$, $\hat{\sigma}_- = 0.14$; positive component: $\hat{\nu}_{+1} = 1.68$, $\hat{\nu}_{+2} = 1.34$, $\hat{\nu}_{+3} = 1.24$, $\hat{\nu}_{+4} = 1.13$, $\hat{\tau}_{+1} = 1.64$, $\hat{\tau}_{+2} = 0.89$, $\hat{\tau}_{+3} = 0.87$, $\hat{\tau}_{+4} = 0.76$, $\hat{\lambda}_{+1} = -11.06$, $\hat{\lambda}_{+2} = -2.47$, $\hat{\lambda}_{+3} = -2.24$, $\hat{\lambda}_{+4} = -1.50$; prevalence estimates: $\hat{\pi}_{+1} = 0.28$, $\hat{\pi}_{+2} = 0.68$, $\hat{\pi}_{+3} = 0.93$, $\hat{\pi}_{+4} = 0.98$). The underlying component that describes the positive results (past infections) is clearly skewed to the left for all age groups.

Figure 8.14 Observed data and fitted mixture model for Slovakia standardised serosurvey results



The residuals are shown below in Figure 8.15. There were some residuals outside the ± 2 range for age group 1-4, which suggests that there were some differences between the observed data and the fitted distributions for the negative underlying component (measurements less than -2 on the \log_{10} -scale).

Figure 8.15 Pearson residuals for Slovakian VZV serosurvey plotted against \log_{10} IU/ml by age group



A comparison between population seroprevalence estimates produced by different methods is shown in Table 8.3.

In these comparisons the results using the mixture model are regarded as the gold standard. Relative to these, the method using the Slovakian assay cut-off underestimate seroprevalences, whereas the method using the Spanish assay cut-off (used in ESEN2) overestimates the seroprevalences. The iterative method gave seroprevalence estimates closer to the “gold standard” mixture model method except for the youngest age group where the Spanish assay cut-off gave a slightly less biased estimate.

Table 8.3 Comparison of seroprevalence estimates between the different methods
(95% CIs in brackets)

Methods	age groups			
	1-4 yrs (n = 456)	5-9 yrs (n = 538)	10-19 yrs (n = 1055)	20+ yrs (n = 1551)
mixture modelling	28.1%	68.2%	92.9%	97.9%
(gold standard)	(23.6%, 33.1%)	(64.0%, 72.1%)	(91.1%, 94.4%)	(97.0%, 98.6%)
Slovakian assay cut-off	19.3%	64.9%	87.4%	93.1%
(applicable to non-standardised results)	(15.8%, 23.2%)*	(60.7%, 68.9%)*	(85.2%, 89.3%)*	(91.7%, 94.3%)*
Spanish ref assay cut-off	32.9%	69.9%	93.2%	97.9%
(used in ESEN2 project)	(28.6%, 37.4%)*	(65.8%, 73.7%)*	(91.5%, 94.6%)*	(97.1%, 98.6%)*
mixture model cut-off (using iterative	22.2%	67.3%	92.0%	97.9%
method - equal misclassification)	(18.4%, 26.2%)*	(63.1%, 71.2%)*	(90.2%, 93.6%)*	(97.1%, 98.6%)*
(age-specific)				

* 95% Binomial exact C

8.9 Method comparison using simulations

A number of simulations were undertaken aiming to compare seroprevalence estimates after applying (a) the reference centre’s cut-off (as used in the ESEN2 project) and (b) the iterative process described in Section 8.7.

The simulation process was as follows:

1. The reference centre’s serosurvey was assumed to follow a mixture distribution, with two underlying components representing the susceptibles (negative group) and the protected individuals (positive group). Location, dispersion and seroprevalence parameters were selected to define this model. The reference centre’s cut-offs were estimated using Equation 8.5.
2. A second mixture model of the same type but with different parameter values was selected for the testing laboratory (true model), and the true model cut-offs were estimated

using Equation 8.5. One thousand sample datasets were generated from this mixture model.

3. For each simulated dataset, seroprevalences were estimated (a) by using the reference centre’s cut-offs estimated in Step 1 and (b) by estimating the cut-offs by the iterative method based on the location and dispersion parameters derived from the reference centre’s model defined in Step 1.

Normal underlying distributions were assumed for the mixture models used in the simulations examples. These were of the general form shown in Equation 7.15.

Simulations – Part 1

The models initially defined for the reference centre and the testing laboratory were assumed to have the same underlying distributions, but different seroprevalences for each age group. These are given in Table 8.4 below.

Table 8.4 Definition of mixture models for testing laboratory (true model) and reference centre (Simulations – Part 1)

(a) Reference centre

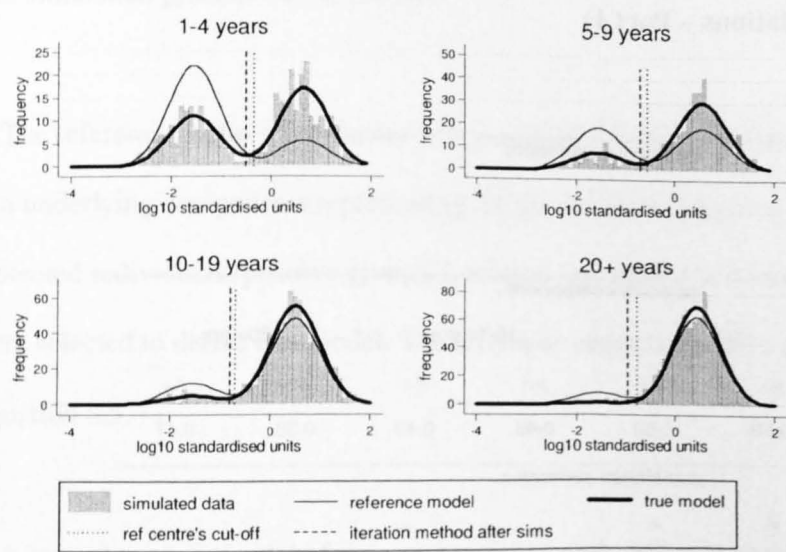
negative component							
μ	σ						
-1.60	0.50						
positive component							
1-4 yrs		5-9 yrs		10-19 yrs		20+ yrs	
μ_{+1}	σ_{+1}	μ_{+2}	σ_{+2}	μ_{+3}	σ_{+3}	μ_{+4}	σ_{+4}
0.60	0.48	0.50	0.50	0.45	0.49	0.38	0.47
prevalence estimates							
π_1	π_2	π_3	π_4				
0.20	0.50	0.80	0.88				

(b) Testing laboratory

negative component							
μ	σ						
-1.60	0.50						
positive component							
1-4 yrs		5-9 yrs		10-19 yrs		20+ yrs	
μ_{+1}	σ_{+1}	μ_{+2}	σ_{+2}	μ_{+3}	σ_{+3}	μ_{+4}	σ_{+4}
0.60	0.48	0.50	0.50	0.45	0.49	0.38	0.47
prevalence estimates							
π_1	π_2	π_3	π_4				
0.60	0.80	0.92	0.98				

The mixture models and the cut-offs are shown in Figure 8.16, together with the mixture models (described in Table 8.4) and the results of one simulation for illustration purposes. Because the testing laboratory had higher seroprevalence than the reference laboratory for all age groups, the cut-off estimated using the iterative method, which takes into account the differences in seroprevalence, was lower than the reference centre cut-off.

Figure 8.16 Comparison between cut-off estimates using the mixture models defined in Table 8.4



Simulations - Part 1

Note: for graphical purposes one set of simulated results was plotted

The age-specific cut-offs were estimated by the different methods following 1,000 simulations. The average cut-off and the seroprevalence estimates obtained by simulation using the different methods are shown in Table 8.5. The cut-offs obtained by the iterative method closely match the cut-offs for the model. There was very little difference between seroprevalence estimates since the underlying distributions were clearly defined and only a small number of samples were around the area where the components meet.

Table 8.5 Comparison between cut-off and seroprevalence estimates methods using the mixture models defined in Table 8.4 (95% percentile intervals in brackets)

models	age groups			
	1-5 yrs	6-9 yrs	10-19 yrs	20+ yrs
<i>prevalence (%)</i>				
true model	0.60	0.80	0.92	0.98
applying true model's cut-off	0.61 (0.56, 0.64)	0.80 (0.78, 0.83)	0.92 (0.90, 0.94)	0.98 (0.97, 0.99)
applying directly ref centre's cut-off	0.59 (0.55, 0.63)	0.79 (0.77, 0.82)	0.91 (0.90, 0.93)	0.97 (0.97, 0.98)
iterative method	0.61 (0.56, 0.65)	0.80 (0.78, 0.84)	0.92 (0.90, 0.94)	0.98 (0.97, 0.99)
<i>cut-off</i>				
true model	-0.52	-0.69	-0.81	-0.97
ref centre's cut-off	-0.35	-0.55	-0.71	-0.78
iterative method	-0.52 (-0.53, -0.50)	-0.69 (-0.71, -0.68)	-0.81 (-0.84, -0.79)	-0.97 (-1.02, -0.93)

Simulations – Part 2

The simulations were repeated for a scenario with a less clear separation between the underlying distributions. The models defined for the reference centre and the testing laboratory are given in Table 8.6 below.

Table 8.6 Definition of mixture models for testing laboratory (true model) and reference centre – overlapping underlying distributions

(a) Reference centre

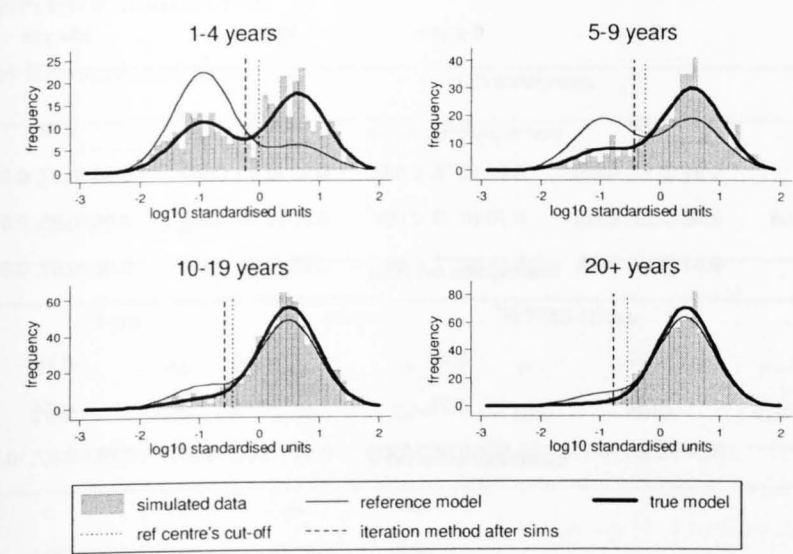
negative component							
μ	σ						
-1.00	0.50						
positive component							
1-4 yrs		5-9 yrs		10-19 yrs		20+ yrs	
μ_{+1}	σ_{+1}	μ_{+2}	σ_{+2}	μ_{+3}	σ_{+3}	μ_{+4}	σ_{+4}
0.60	0.48	0.50	0.50	0.45	0.49	0.38	0.47
prevalence estimates							
π_1	π_2	π_3	π_4				
0.20	0.50	0.80	0.88				

(b) Testing laboratory

negative component							
μ	σ						
-1.00	0.50						
positive component							
1-4 yrs		5-9 yrs		10-19 yrs		20+ yrs	
μ_{+1}	σ_{+1}	μ_{+2}	σ_{+2}	μ_{+3}	σ_{+3}	μ_{+4}	σ_{+4}
0.60	0.48	0.50	0.50	0.45	0.49	0.38	0.47
prevalence estimates							
π_1	π_2	π_3	π_4				
0.60	0.80	0.92	0.98				

The mixture models and the estimated cut-offs obtained through the different methods are shown in Figure 8.17. The underlying components are close together, however, there are clear peaks that are especially visible for the younger age groups which include higher numbers of negative samples.

Figure 8.17 Comparison between cut-off estimates using the mixture models defined in Table 8.6



Simulations - Part 2

Note: for graphical purposes one set of simulated results was plotted

The simulated cut-offs from 1,000 runs and the seroprevalence estimates obtained with the different methods are given in Table 8.7. Again the iterative method gave cut-off estimates close to the truth, whereas the reference centre's cut-offs were underestimates. The seroprevalences obtained using the iterative method are close to those of the true model, as in the previous example. In contrast, direct application of the estimated reference age-specific cut-offs underestimates the true seroprevalence. The reason for this is that there is higher number of observations around the area where the underlying component distributions intersect which affects the sample classification.

Table 8.7 Comparison between cut-off and seroprevalence estimates methods using the mixture models defined in Table 8.6 (95% percentile intervals in brackets)

models	age groups			
	1-5 yrs	6-9 yrs	10-19 yrs	20+ yrs
<i>prevalence (%)</i>				
true model	0.60	0.80	0.92	0.98
applying true model's cut-off	0.61 (0.57, 0.65)	0.80 (0.78, 0.83)	0.92 (0.90, 0.94)	0.98 (0.97, 0.99)
applying directly ref centre's cut-off	0.56 (0.52, 0.60)	0.76 (0.73, 0.80)	0.90 (0.88, 0.92)	0.96 (0.95, 0.97)
iterative method	0.61 (0.56, 0.65)	0.81 (0.77, 0.84)	0.92 (0.90, 0.94)	0.98 (0.97, 0.99)
<i>cut-off</i>				
true model	-0.23	-0.43	-0.58	-0.78
ref centre's cut-off	-0.02	-0.25	-0.45	-0.54
iterative method	-0.24 (-0.26, -0.21)	-0.43 (-0.46, -0.41)	-0.58 (-0.62, -0.55)	-0.78 (-0.87, -0.72)

Simulations – Part 3

The simulations were repeated once more, after violating the iterative method assumption i.e. that the underlying component distributions are the same, and hence have the same location and dispersion parameters. The mixture models defined for the reference centre and the testing laboratory are given in Table 8.8 below.

Table 8.8 Definition of mixture models for testing laboratory (true model) and reference centre – varying location and dispersion parameters between testing and reference laboratories

(a) Reference centre

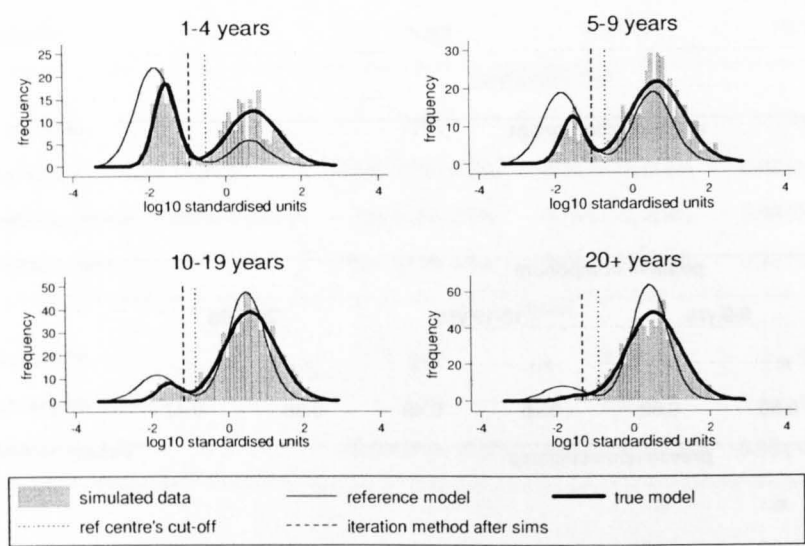
negative component							
μ	σ						
-1.90	0.50						
positive component							
1-4 yrs		5-9 yrs		10-19 yrs		20+ yrs	
μ_{+1}	σ_{+1}	μ_{+2}	σ_{+2}	μ_{+3}	σ_{+3}	μ_{+4}	σ_{+4}
0.60	0.48	0.50	0.50	0.45	0.49	0.38	0.47
prevalence estimates							
π_1	π_2	π_3	π_4				
0.20	0.50	0.80	0.88				

(b) Testing laboratory

negative component							
μ_{-}	σ_{-}						
-1.60	0.30						
positive component							
1-4 yrs		5-9 yrs		10-19 yrs		20+ yrs	
μ_{+1}	σ_{+1}	μ_{+2}	σ_{+2}	μ_{+3}	σ_{+3}	μ_{+4}	σ_{+4}
0.70	0.68	0.60	0.70	0.55	0.69	0.48	0.69
prevalence estimates							
π_1	π_2	π_3	π_4				
0.60	0.80	0.92	0.98				

The mixture models and estimated cut-offs are given in Figure 8.18. There is a clear difference between the reference centre and the testing laboratory’s models for both underlying components.

Figure 8.18 Comparison between cut-off estimates using the mixture models defined in Table 8.14



Simulations - Part 3

Note: for graphical purposes one set of simulated results was plotted

The simulated cut-offs (from 1000 runs) and seroprevalence estimates obtained by different methods are given in Table 8.9. Examining the iterative method cut-off estimates, it is clear that these are not as close to the true model's cut-off as for the simulation examples shown before. Still, however, they are better than the cut-offs estimated using the reference centre's model. The differences in seroprevalence between the two methods were small, due to the fact that there were not many samples around the area where the underlying component distributions meet.

Table 8.9 Comparison between cut-off and seroprevalence estimates methods using the mixture models defined in Table 8.8 (95% percentile intervals in brackets)

models	age groups			
	1-5 yrs	6-9 yrs	10-19 yrs	20+ yrs
<i>prevalence (%)</i>				
true model	0.60	0.80	0.92	0.98
applying true model's cut-off	0.61 (0.57, 0.65)	0.80 (0.78, 0.83)	0.92 (0.90, 0.94)	0.98 (0.97, 0.99)
applying directly ref centre's cut-off	0.58 (0.54, 0.62)	0.78 (0.75, 0.81)	0.90 (0.88, 0.92)	0.96 (0.95, 0.97)
iterative method	0.59 (0.55, 0.63)	0.79 (0.76, 0.82)	0.90 (0.89, 0.92)	0.97 (0.95, 0.98)
<i>cut-off</i>				
true model	-0.93	-1.06	-1.17	-1.35
ref centre's cut-off	-0.51	-0.70	-0.84	-0.90
iterative method	-0.66 (-0.67, -0.64)	-0.82 (-0.84, -0.81)	-0.92 (-0.94, -0.90)	-1.03 (-1.07, -1.00)

8.10 Advantages and disadvantages of each method

The method used for the ESEN2 project has the advantage of transforming serological results from different sources into comparable units, making it possible to carry out different types of analysis. One disadvantage is its great dependency on the (reference centre) assay manufacturer’s cut-off point, which was constructed solely for diagnostic purposes.

ROC analysis can be used to obtain cut-offs for national serosurveys by comparing each country’s results to the reference serosurvey. The advantage of this method is that it is easy to apply and avoids fitting regression equations. However, it is only expected to work for assays with a clear distinction between positive and negative results.

Reliable seroprevalence estimates can be achieved by the use of mixture modelling on each laboratory’s results. However, abandoning standardisation in favour of such an approach would mean that no direct comparisons would be possible between quantitative serological results. Moreover, in order to obtain unbiased seroprevalence estimates using mixture modelling, an appropriate mixture model must be selected. This is not always easy or even

possible, especially when underlying distributions are not well separated (see previous Chapter).

In this Chapter, a combination of standardisation and mixture model methods has been proposed for estimating seroprevalence. Specifically, mixture modelling may be used to re-estimate a cut-off for the reference centre, and subsequently, this may apply to the other standardised serosurveys in an iterative process. The idea was to compare the two methods, and in particular the method based on standardised results, with that based on cut-offs that are more appropriate for seroprevalence studies. The disadvantage of the iterative method is its dependence on the assumption that the components should be identical between serosurveys. While this should be addressed to some extent by standardisation, some variation may be expected owing to differences between the populations due to other factors.

8.11 Conclusions

The seroprevalence estimation method used during ESEN2 was based on assay cut-offs. The aim of this Chapter was to propose techniques for re-estimating assay cut-off for sero-epidemiology. The iterative method described above qualifies as an alternative to the assay cut-off method, especially when (a) there is no widely accepted cut-off or the reference centre's cut-off is clearly not appropriate, and (b) it is not possible to use mixture models as an alternative due to non-separable underlying distributions. For the VZV example used, an established cut-off existed for the Spanish serosurvey, but mixture models could also be fitted given the well-separated distributions between past infections and uninfected individuals. This particular example was selected to present the different methods and compare the estimated seroprevalences.

In the example shown in this Chapter, the method used for the ESEN2 project provided robust for all ages apart for the youngest age group, and the resulting estimates were comparable to those obtained with the iterative method. This provides re-assurance with regards to the suitability of the manufacturer's cut-off, in spite of its lack of motivation for the purpose of estimating seroprevalence. The iterative method proposed may be used as a suitable alternative in a future ESEN project.

The key assumption for the iterative method to be successful is that the underlying components between the reference centre and the testing laboratory are the same (and hence, have similar location and dispersion parameters), given the serological data have been standardised. As was shown in the last simulation example, varying the parameter estimates affects the estimated cut-offs, although in that particular scenario it did not much alter the seroprevalence estimates.

Finally, despite the fact that mixture models often provide better seroprevalence estimates than the fixed cut-offs methods, there are circumstances where such cut-offs may be useful. For example, once fixed cut-offs have been specified, they can be easily applied to other serosurvey studies, assuming there is no change in the serological methods. This means that more reliable seroprevalence estimates can be obtained from smaller studies than those that would need to apply mixture modelling techniques.

Chapter 9: Seroprevalence estimation taking into account plate-to-plate variability

9.1 Introduction

In the previous Chapters various methods for obtaining comparable population seroprevalence estimates have been presented. Using these methods it is possible to adjust for sources of variability such as those between different laboratories or assays. However, a certain amount of variability would still be present even if samples were tested using a unique assay within the same laboratory.

Some of the unexplained variability may be attributed to the fact that not all samples are tested simultaneously. Instead batches are placed and tested on different assay plates. For many of these assays, the manufacturers adjust for such variability by transforming the optical density (OD) results from each plate into International Units (IU/ml) (Chapter 2).

The aim of this Chapter is to investigate the impact of variability due to plate differences on the seroprevalence estimate comparisons, and to propose a method to adjust for this.

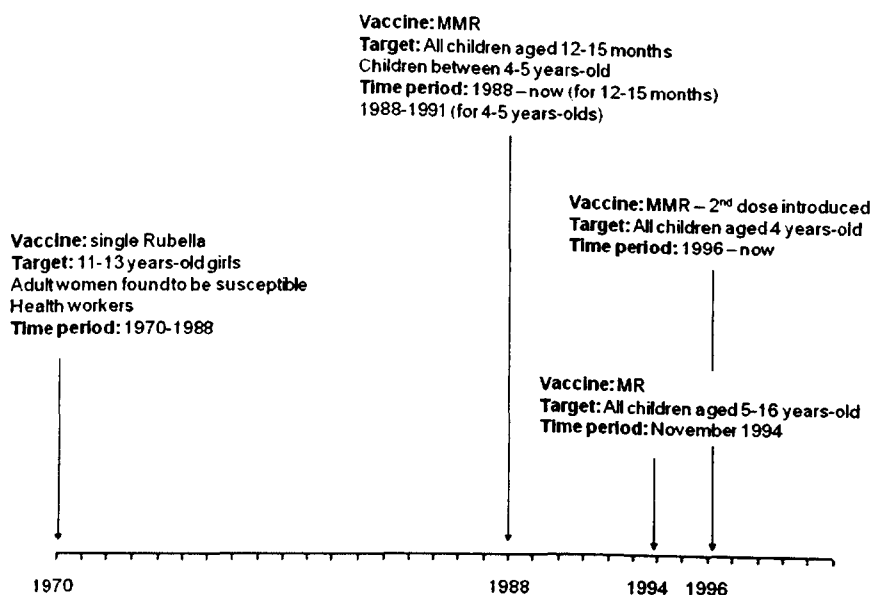
9.2 Rubella sero-survey

The UK rubella sero-survey example, which will be used throughout this Chapter, was tested in 2000 and comprised 3,460 samples. A Mercia Rubella-G test was employed (Microgen Bioproducts Ltd, Camberley, UK) to classify individual samples into positive and negative using a cut-off at 8 IU/ml. Sera from individuals aged from 1 to 69 years of age were included (Vyse AJ *et al.*, 2006). For the purposes of the analysis the data were divided into 4 different age groups: 1-9, 10-17, 18-43 and 44+. To understand the reasons for selecting these particular age groups it is essential to briefly describe the history of rubella vaccination in the UK.

The single rubella vaccine was introduced in the UK in 1970, targeting schoolgirls aged 11-13 years-old, susceptible women (mainly identified through antenatal testing) and

healthcare workers. The combined vaccine against measles, mumps and rubella (MMR) was introduced during 1988 for all children aged 12-15 months. It was also offered to children aged between 4 and 5 for 3 years until 1991, administered at the same time as the pre-school doses of the diphtheria / tetanus vaccine. The idea was for children of these ages to be protected until the first cohort that received the first dose at 12-15 months reached the age of 4 years (Miller E, 1991). Due to a rubella outbreak in 1993, a “catch-up” campaign was introduced in November 1994. The combined measles and rubella vaccine was then offered to all children between 5 and 16 years-old. Since 1996 the MMR vaccine is given in 2 doses, the second administered to 3-5 years-olds (Vyse AJ *et al.*, 2002;Nardone A *et al.*, 2008;Tookey P, 2004;Hardelid P, 2008). Figure 9.1 shows the history of the rubella vaccination strategy in the UK.

Figure 9.1 Rubella vaccination programme for the UK



Using this information, the rationale for grouping the data from the ESEN2 serosurveys into 4 broad age groups was the following:

- (a) Individuals who were older than 13 years in 1970 did not receive any vaccine (i.e. at least 44 years old in 2000 when the sero-survey took place).
- (b) Those older than 5 years in 1988 (and younger than 13 years in 1970) were targeted in the single rubella campaign (these would have been 18-43 in 2000).
- (c) Those older than 4 years in 1996 (and 5 years or younger in 1988) were part of the single MMR or part of the MR catch-up campaign (these would have been 10-17 in 2000).
- (d) Those younger than 9 years old were all part of the 2 dose MMR programme, although those less than 3 had not received the second dose at the time the sero-survey was carried out (these would have been 1-9 years-old in 2000) (Table 9.1).

Table 9.1 Age group classifications according to vaccination status

age groups	birth cohort	Vaccination status
1-9	> 1991	two MMR doses*
10-17	1983-1990	single MMR or MR dose
18-43	1957-1982	single rubella vaccine
44+	< 1956	no vaccination

* except for those who were too young to have received the 2nd dose (2nd dose given between 3-5 years)

Although gender differences are important, since the single rubella programme was initially aimed mainly at women, unfortunately, gender information was not available and therefore, could not be taken into account in the analysis.

9.3 Optical Density (OD) measurements

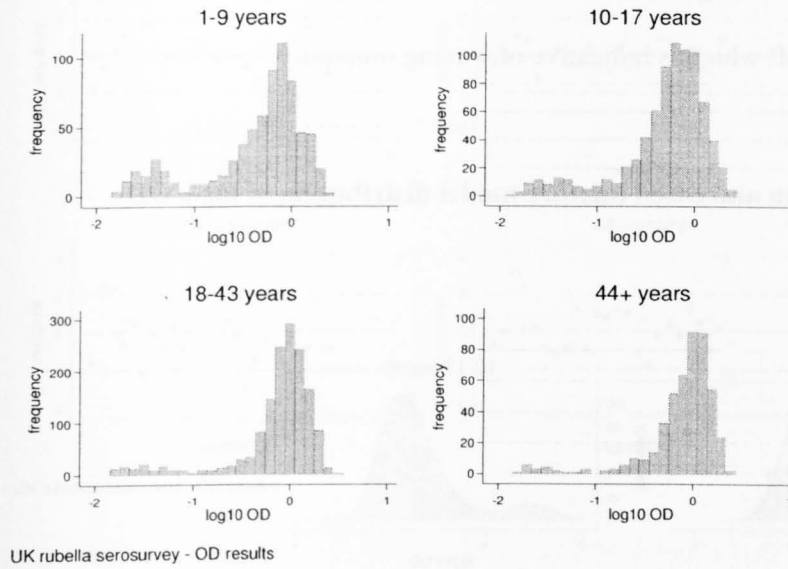
As was explained in Chapter 2, the samples from each population serosurvey are placed on different plates. In this research each plate had 12 columns and 8 rows of wells, and therefore, a maximum of 96 samples could be tested (Figure 2.4 and Figure 9.2).

Figure 9.2 The design of the serological plates used for the rubella tests by the UK laboratory

A	1	2	3	4	5	6	7	8	9	10	11	12
B												
C												
D												
E												
F												
G												
H												

After each plate was tested, the outcomes were reported quantitatively as ODs (Chapter 2). For the UK rubella serosurvey, the frequency distribution of the log-transformed OD measurements by each age group is shown in Figure 9.3. The distribution consists of two peaks that denote the underlying groups of negative and positive results.

Figure 9.3 Frequency distribution of \log_{10} OD measurements for UK rubella serosurvey



The underlying distribution of the positive results would be expected to include both past infections and vaccinated individuals, although the antibody response resulting from vaccination may be lower than that arising from infection by the wild virus. Assuming a normal distribution for the negative results and a skew normal for the positive, a mixture model of the type expressed in Equation 7.16 was fitted with age groups 1-9, 10-17, 18-43 and more than 43 years.

The resulting estimates were:

Negative component: $\hat{\mu}_- = -1.48$, $\hat{\sigma}_- = 0.23$.

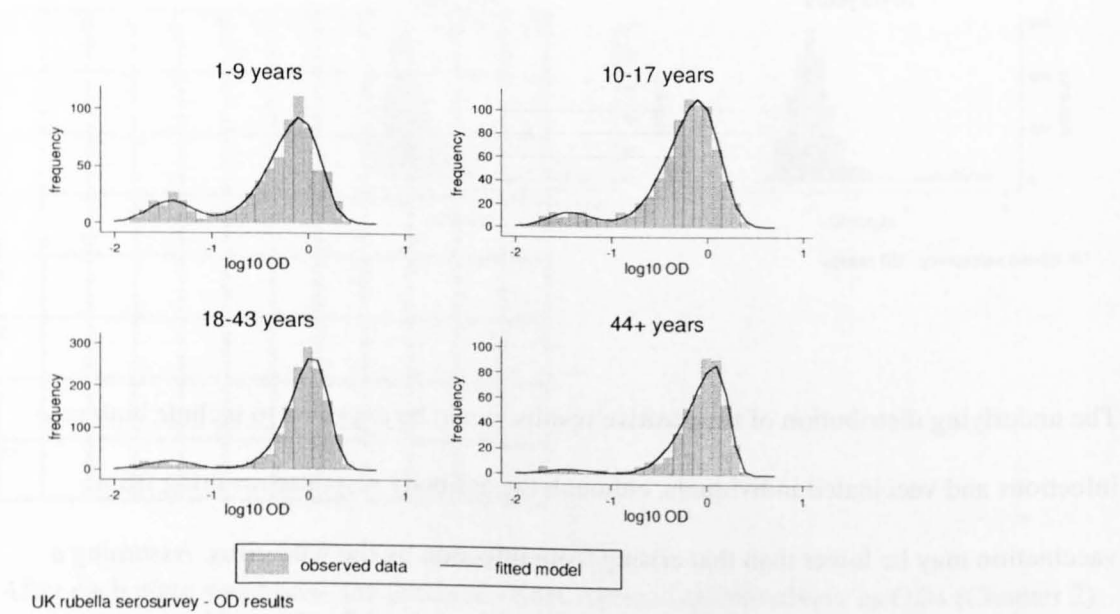
Positive component: $\hat{v}_{+1} = 0.07$, $\hat{v}_{+2} = 0.06$, $\hat{v}_{+3} = 0.15$, $\hat{v}_{+4} = 0.18$, $\hat{\tau}_{+1} = 0.41$, $\hat{\tau}_{+2} = 0.40$,

$\hat{\tau}_{+3} = 0.32$, $\hat{\tau}_{+4} = 0.36$, $\hat{\lambda}_1 = -2.57$, $\hat{\lambda}_2 = -2.32$, $\hat{\lambda}_3 = -2.69$, $\hat{\lambda}_4 = -4.17$.

Prevalence estimates: $\hat{\pi}_1 = 0.84$, $\hat{\pi}_2 = 0.91$, $\hat{\pi}_3 = 0.92$, $\hat{\pi}_4 = 0.96$.

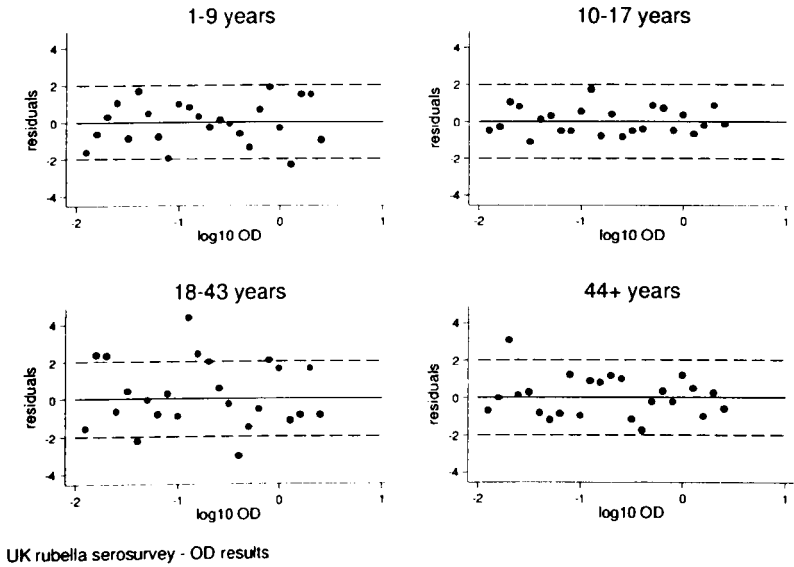
The fitted model is plotted in Figure 9.4 below. The underlying component of the positive results is skewed to the left which is indicative of waning immunity.

Figure 9.4 Observed data and fitted mixture model distribution of \log_{10} OD measurements



The Pearson residual plot is given in Figure 9.5. The residuals are generally within the ± 2 limits for all age groups apart from ages 18 to 43 years, where there were differences between the observed data and the fitted distribution between the range of -0.9 and 0.1 on the \log_{10} OD scale.

Figure 9.5 Pearson residuals by age group for log₁₀ OD measurements



The drawback of OD measurements is that they are not comparable between plates. Any slight change in the conditions in which a plate is tested or in the individual carrying out the test might affect the outcomes. The variation in OD measurements between the plates will be referred to as *plate-to-plate* variability.

9.4 Control samples

Assay manufacturers are aware that OD is not a reliable measurement owing to plate-to-plate variability. A simple way to adjust for this is to include a small quantity of a unique negative sample in each plate to be tested to act as a *control sample* (Vyse AJ *et al.*, 2006). *Adjusted OD measurements* OD_A , may then be calculated by using the following equation:

$$OD_{Aij} = \frac{OD_{ij}}{OD_{Cj}},$$

Equation 9.1

where OD_{ij} refers to the OD result from the i^{th} sample on the j^{th} plate and OD_{Cj} is the OD measurement of the control sample on the j^{th} plate. The adjusted OD is often known as “test to negative control ratio” (T/N) (Vyse AJ *et al.*, 2006).

Assuming, as before, a normal distribution for the negative underlying component and a skew normal for the positive, a mixture model of the type shown in Equation 7.16 was fitted for the adjusted OD measurements. The resulting estimates were:

Negative component: $\hat{\mu}_- = 0.28$, $\hat{\sigma}_- = 0.24$.

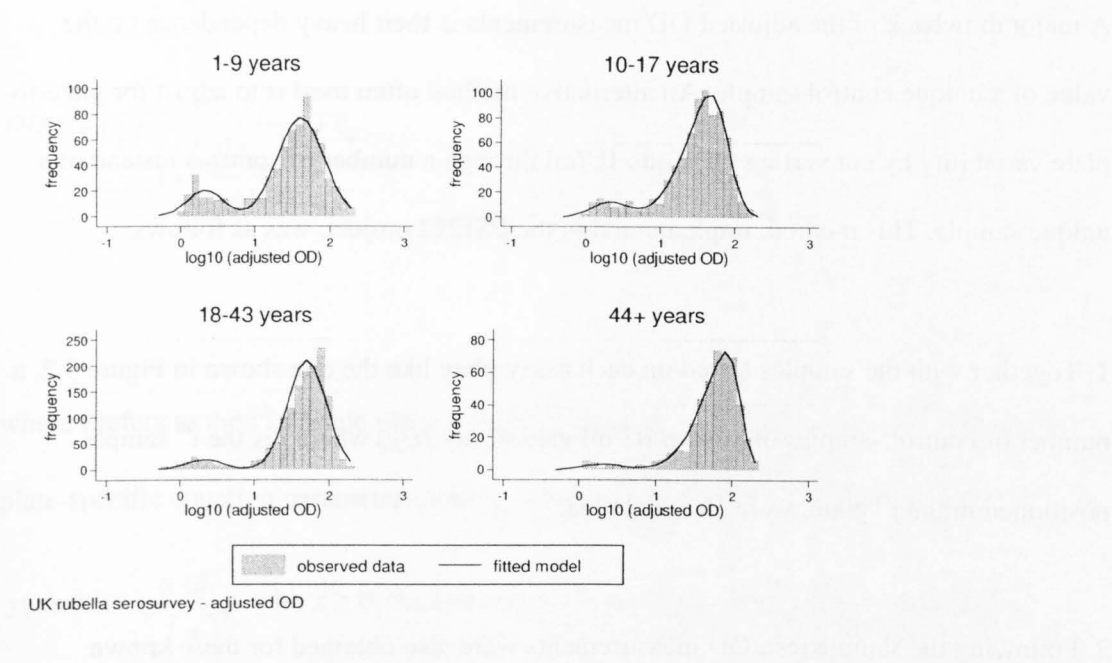
Positive component: $\hat{v}_{+1} = 1.56$, $\hat{v}_{+2} = 1.84$, $\hat{v}_{+3} = 1.66$, $\hat{v}_{+4} = 2.05$, $\hat{\tau}_{+1} = 0.31$, $\hat{\tau}_{+2} = 0.45$,
 $\hat{\tau}_{+3} = 0.26$, $\hat{\tau}_{+4} = 0.42$, $\hat{\lambda}_{+1} = 0.01$, $\hat{\lambda}_{+2} = -2.41$, $\hat{\lambda}_{+3} = 0.00$, $\hat{\lambda}_{+4} = -3.39$.

Prevalence estimates: $\hat{\pi}_1 = 0.83$, $\hat{\pi}_2 = 0.91$, $\hat{\pi}_3 = 0.93$, $\hat{\pi}_4 = 0.96$.

The frequency distribution and the resulting fitted model are shown in Figure 9.6.

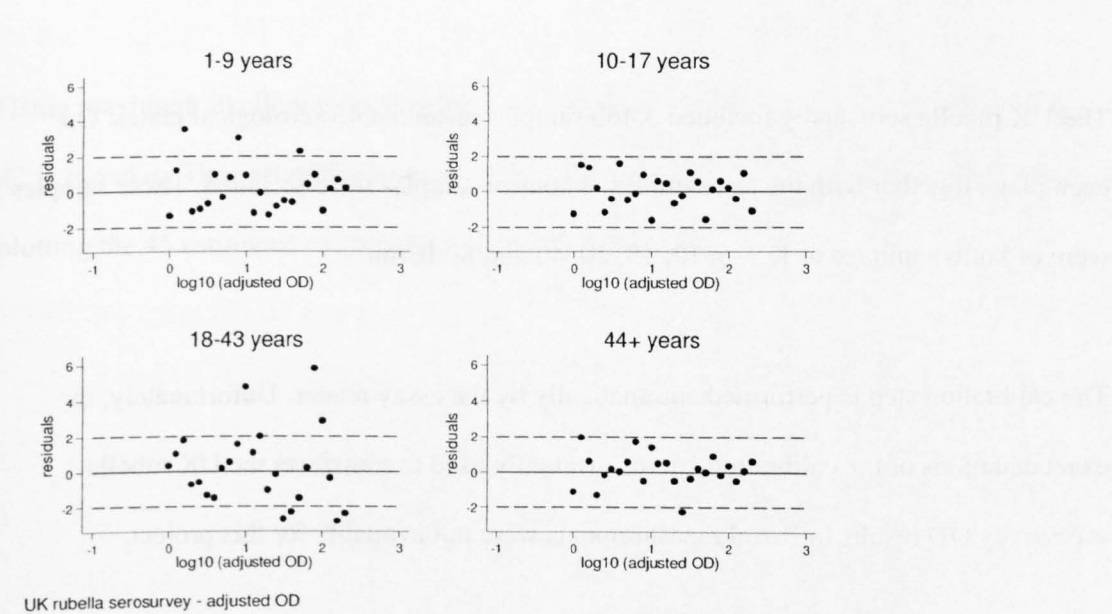
Although the scale measurements are naturally different in comparison to the crude ODs, the shape of the distributions is similar.

Figure 9.6 Observed data and fitted mixture model distribution of $\log_{10} \text{OD}_A$ measurements



The residual plots are given in Figure 9.7 below. There are differences between the observed data and the fitted distribution for the positive underlying component for ages between 18 and 43 years.

Figure 9.7 Pearson residuals by age group for $\log_{10} \text{OD}_A$ measurements



9.5 Calibration curves

A major drawback of the adjusted OD measurements is their heavy dependence on the value of a unique control sample. An alternative method often used is to adjust for plate-to-plate variability by converting ODs into IU/ml through a number of controls instead of a unique sample. This method, implemented in the ESEN2 project, was as follows:

1. Together with the samples tested on each assay plate like the one shown in Figure 9.2, a number of control samples of known IU/ml values, say IU_{Kij} where i is the i^{th} sample positioned in the j^{th} plate, were also included.
2. Following the sample test, OD measurements were also obtained for these known control samples, say OD_{Lij} .
3. After regressing IU_{Kij} against OD_{Lij} , plate-specific calibration curves were estimated.
4. The calibration curves from (3) were used to transform the samples' OD results in each plate (OD_{ij}) into IU/ml (IU_{ij}).

The UK rubella serosurvey included 3,460 samples tested on 45 serological plates. On each plate, together with the test samples, 8 control samples were included. These samples were of known unitage at 1, 2, 5, 10, 15, 20, 40 and 80 IU/ml.

The calibration step is performed automatically by the assay reader. Unfortunately, the exact equations of the calibration curves originally used to transform the UK rubella serosurvey OD results to IU/ml measurements were not available for this project.

Therefore, as an example, an attempt to reconstruct the process of generating IU/ml from OD measurements was carried out using a calibration curve of the following type:

$$OD_{Lij} = \delta_j + \frac{\alpha_j - \delta_j}{1 + \left(\frac{IU_{Kij}}{\gamma_j} \right)^\beta} + \varepsilon_{ij},$$

Equation 9.2

where i refers to the i^{th} sample placed on the j^{th} plate ($j = 1, \dots, 45$) and $\alpha_j, \beta_j, \gamma_j$ and δ_j are plate-specific equation parameters, and $\varepsilon_{ij} \sim N(0, \sigma^2)$. Note that for the function

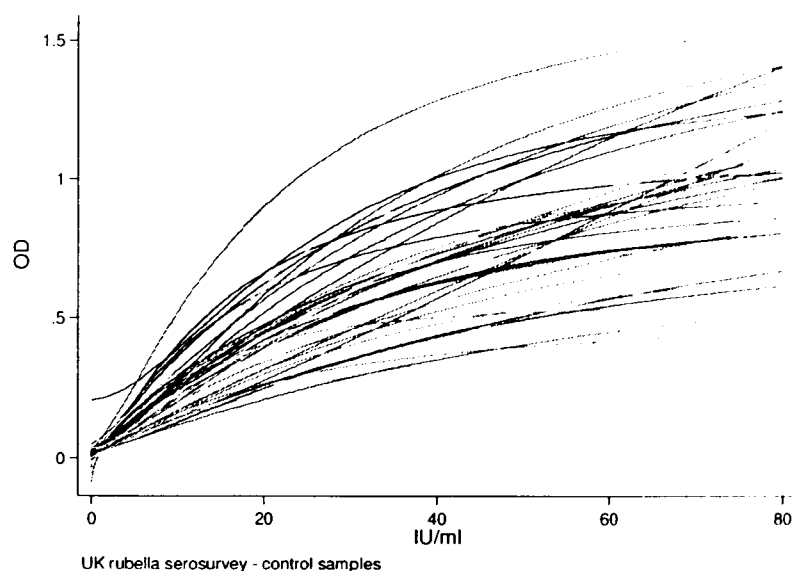
$$y(x) = \delta + \frac{\alpha - \delta}{1 + \left(\frac{x}{\gamma} \right)^\beta} \text{ with } x \geq 0, \text{ the lower horizontal asymptote is } y(0) = \alpha, \text{ the upper}$$

$\lim_{x \rightarrow +\infty} (y(x)) = \delta$, whereas the parameters β and γ correspond to the slope and location,

respectively. The calibration curve shown has been used to transform OD results into IU/ml for another assay than the one shown in this example. Although different curves are used by different assays these are often of similar types so it was hoped that the impact on the outcome would be limited.

Using maximum likelihood for Equation 9.2, the following parameters were estimated: $\alpha_j, \beta_j, \gamma_j, \delta_j$ and σ_j . The variability between calibration equation estimates may be viewed by plotting the 45 estimated calibration equations (Figure 9.8).

Figure 9.8 Generated calibration curves for the 45 plates



Since the calibration equations were estimated using control samples with known unitage, it was possible to obtain IU/ml measurements for the rest of the samples on the plates by inverting Equation 9.2 as shown below:

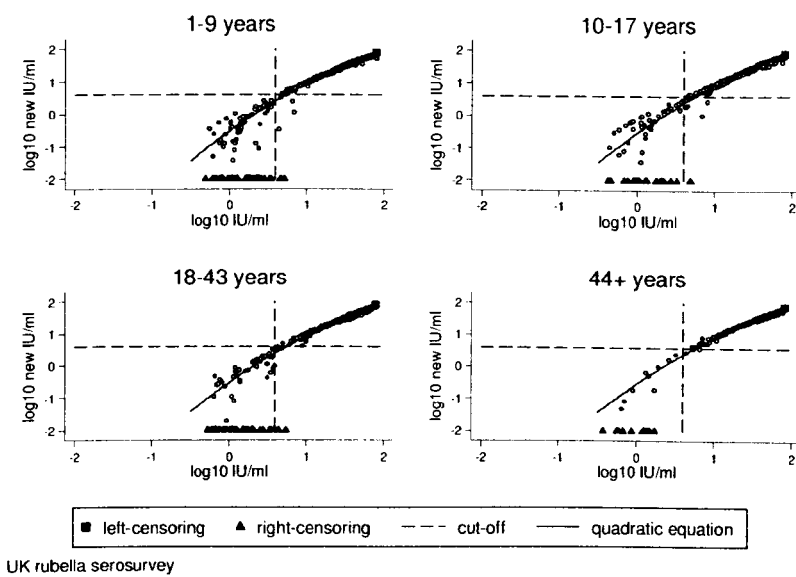
$$IU_{ij} = \hat{\gamma}_j \left(\frac{\hat{\alpha}_j - \hat{\delta}_j}{OD_{ij} - \hat{\delta}_j} - 1 \right)^{\frac{1}{\hat{\beta}}}$$

Equation 9.3

A comparison between the IU/ml generated automatically by the assay reader (x-axis) and those estimated by the Equations 9.2 and 9.3 (y-axis) is given in Figure 9.9 after these were log₁₀-transformed. The samples measured outside the range 0.01 to 80 IU/ml were treated as censored and were replaced with “0.01 IU/ml” or “80 IU/ml” if they were below or above the range, respectively. The assay cut-off, given as 4 IU/ml (or 0.6 on the log₁₀-scale), is also shown in Figure 9.9. A quadratic equation was fitted to the quantitative results, aiming to describe the relationship between the two sets of results (the quadratic

term was tested significant $p < 0.001$). In general, the fitted lines explain well the data (coefficient of determination $R^2 = 0.97$), however, there is clearly larger variability for lower readings.

Figure 9.9 Comparison between IU/ml given by the assay reader (\log_{10} IU/ml) and as estimated using the generated calibration curves (\log_{10} new IU/ml)



The conversion of OD measurements into IU/ml shown in this Section can only serve as an example as to how the transformation had been implemented, given that the actual calibration equation used was unknown. From this point onward, the IU/ml measurements provided by the UK laboratory will be used for the analysis.

9.6 Estimating plate-to-plate variability using the control samples

As mentioned above, each plate had a number of control samples of known IU/ml value, say IU_{Kij} , from which an OD measurement was obtained following the test (say OD_{Lij}). Following the estimation of the calibration equations, the OD_{Lij} measurements were transformed to IU/ml, say IU_{Lij} for the i^{th} control sample placed in the j^{th} plate.

The variability between plates was investigated using the control samples. A one-way analysis of variance (ANOVA) was performed using the model below.

$$y_{ij} = \mu + \alpha_j + \varepsilon_{ij} ,$$

Equation 9.4

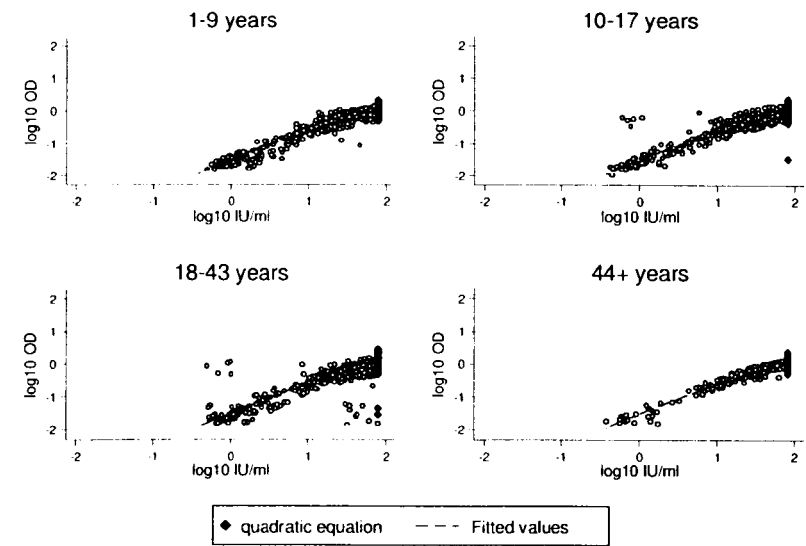
where μ stands for the total mean, α_j is the deviation from the total mean for the j^{th} plate ($j = 1, \dots, \psi$) and ε_{ij} the error term assumed to be identically and independently normally distributed with zero mean and constant variance σ^2 i.e. $\varepsilon_{ij} \sim N(0, \sigma^2)$ (Zar JH, 1999).

For the 357 control samples placed on 45 plates (three controls were missing from one plate), the one-way ANOVA using $\log_{10} IU_{Lij}$ measurements as a response variable, showed no significant evidence that the samples were drawn from different populations ($p = 1.00$). The between-plate variability was quantified using the *mean square error (MSE)* which was estimated as 0.02. This was much lower compared to the estimate obtained after using the $\log_{10} OD_{Lij}$ measurements as a response variable ($\hat{MSE} = 0.15$), which indicates that the transformation of OD results to IU/ml corrected for plate-to-plate variability.

9.7 International Units (IU/ml)

A comparison between OD and IU/ml measurements by age group is given in Figure 9.10. A quadratic equation was selected to describe the relationship between the two sets of results given that the quadratic term significantly improved the model fit ($p < 0.001$). Although the proportion of the variance explained by the regression is high (overall $R^2 = 0.85$), there are a number of discrepant observations that have high OD but low IU/ml results (age groups 10-17 and 18-43) or low OD and high IU/ml (age group 18-43).

Figure 9.10 Comparison between \log_{10} OD and \log_{10} IU/ml by age group



UK rubella serosurvey

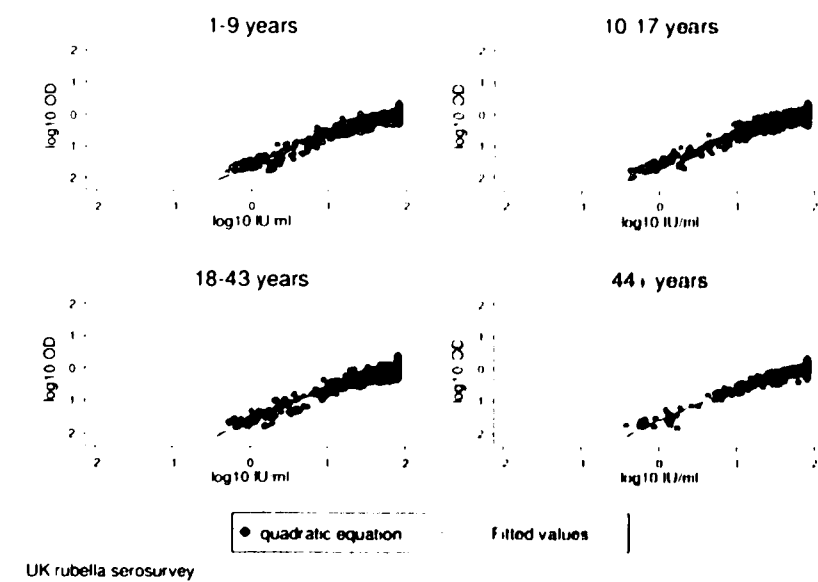
Looking at the extreme observations, it turns out that they all belong to one plate (plate number 38). Table 9.2 shows the true IU/ml of the control samples and the values generated from the machine reading for plate 38. Given that the machine-generated IU/ml measurements (IU_{Li38}) are close to the true IU/ml (IU_{Ki38}), this suggests that the calibration equation has been successfully estimated but not correctly applied to the rest of the samples on the plate.

Table 9.2 Comparison between the true IU/ml, the OD and the machine-generated IU/ml for the control samples of plate 38

IU_{KB8} (IU/ml)	OD_{LB8}	IU_{LB8} (IU/ml)
1	0.043	1.242
2	0.055	1.855
5	0.117	6.36
10	0.128	8.64
15	0.354	9.913
20	0.406	27.415
40	0.727	38.823
80	1.101	81.237

Since the generated IU/ml results for plate 38 were considered to be biased, they were dropped from any further analysis. Figure 9.11 shows the relationship between OD and IU/ml after plate 38 was excluded from the analysis.

Figure 9.11 Comparison between \log_{10} OD and \log_{10} IU/ml by age group'after excluding plate 38



9.8 Mixture modelling using IU/ml

A mixture model was fitted to the machine-generated log₁₀-transformed IU/ml using Equation 7.16, after adjusting to take into account the censored observations as shown in Equation 7.6. The resulting estimates were:

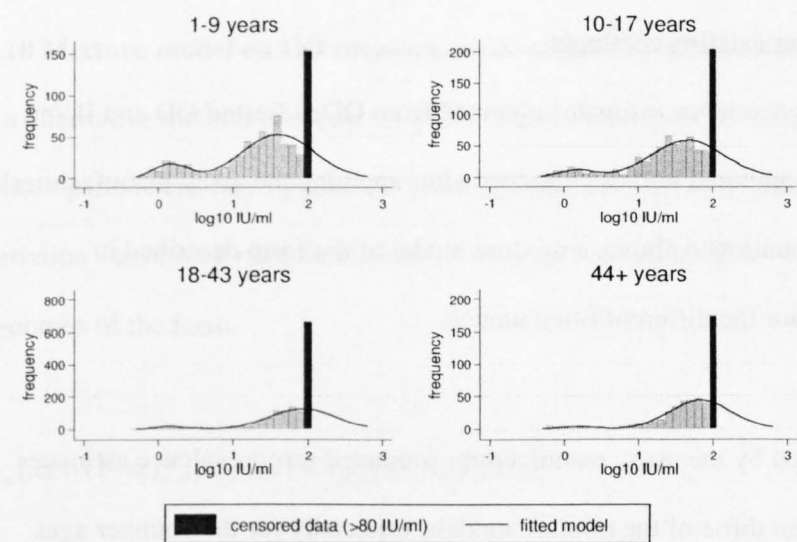
Negative component: $\hat{\mu}_- = 0.11$, $\hat{\sigma}_- = 0.23$.

Positive component: $\hat{v}_{+1} = 1.56$, $\hat{v}_{+2} = 1.61$, $\hat{v}_{+3} = 1.88$, $\hat{v}_{+4} = 1.77$, $\hat{\tau}_{+1} = 0.47$, $\hat{\tau}_{+2} = 0.48$, $\hat{\tau}_{+3} = 0.43$, $\hat{\tau}_{+4} = 0.41$, $\hat{\lambda}_{+1} = -0.76$, $\hat{\lambda}_{+2} = -0.73$, $\hat{\lambda}_{+3} = -0.84$, $\hat{\lambda}_{+4} = -0.88$.

Prevalence estimates: $\hat{\pi}_1 = 0.86$, $\hat{\pi}_2 = 0.91$, $\hat{\pi}_3 = 0.93$, $\hat{\pi}_4 = 0.96$.

The fitted and observed distributions are shown in Figure 9.12. The figure shows that approximately, a third of the samples were measured as “> 80 IU/ml” (right-censored).

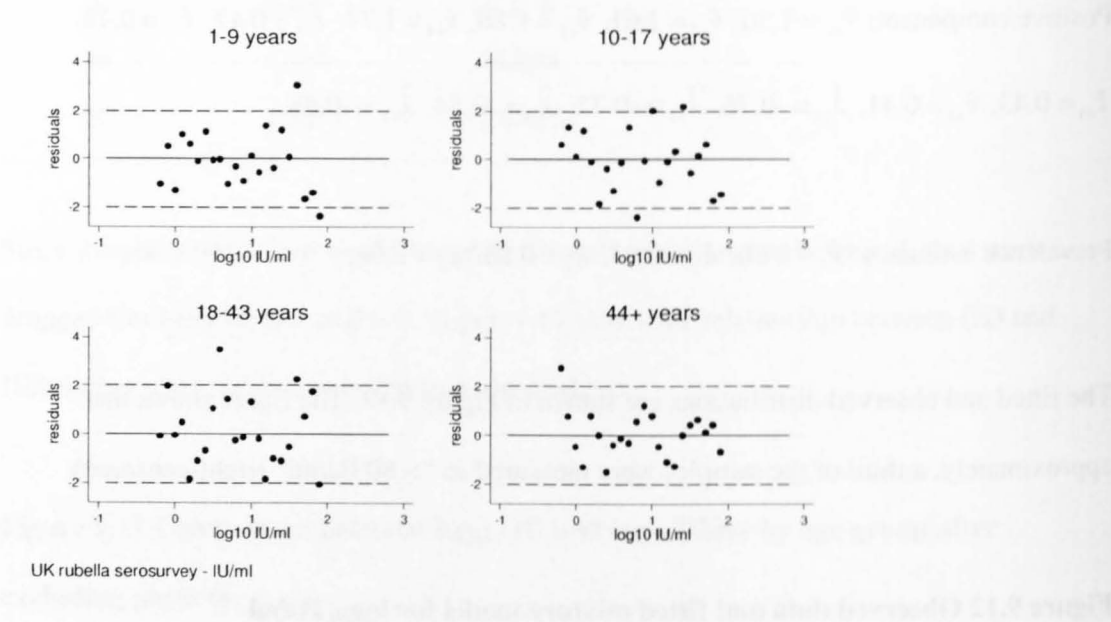
Figure 9.12 Observed data and fitted mixture model for log₁₀ IU/ml



UK rubella serosurvey - IU/ml

The residuals plots are given in Figure 9.13. For age group 1-9 years there are a few measurements at higher IU values that have residuals outside the ± 2 limits, which suggests some lack of fit on the positive underlying component of the model. For age groups 18-43 and more than 44 years, there are a couple of lower measurements with extreme residuals that suggest some lack of fit on the negative component.

Figure 9.13 Pearson residuals by age group for \log_{10} IU/ml



9.9 Comparison between existing methods

The mixture model seroprevalence estimates obtained from OD, adjusted OD and IU/ml measurements were all compared to those obtained after applying the assay manufacturer’s cut-off (Table 9.3). As mentioned above, a mixture model of the form described in Equation 7.16 was used for the different fitted models.

Using the cut-off specified by the assay manufacturer produced seroprevalence estimates that were much lower than those of the mixture models, especially for the younger ages. These differences may be due to the choice of cut-off (as discussed in Chapter 8).

There were no large differences between seroprevalence estimates resulting from the mixture models. However, the seroprevalence estimates generated from IU/ml were slightly higher than those derived from OD or adjusted OD (especially for the younger age groups), which may suggest that this method corrects for slight bias resulting from plate-to-plate variability.

Table 9.3 UK population seroprevalence estimates for rubella given by age groups (95% CI in brackets)

methods	age groups			
	1-9 yrs	10-17 yrs	18-43 yrs	44+ yrs
assay classification (IU/ml)	78.9%	86.1%	91.6%	93.6%
	(75.6%, 81.8%)*	(83.5%, 88.5%)*	(90.1%, 93.0%)*	(91.0%, 95.6%)*
mixture model – OD	84.1%	90.8%	92.4%	95.5%
	(81.0%, 86.7%)	(88.3%, 92.7%)	(90.9%, 93.7%)	(93.2%, 97.1%)
mixture model - adjusted OD	83.2%	90.8%	92.5%	95.7%
	(80.1%, 85.9%)	(88.2%, 92.8%)	(91.0%, 93.8%)	(93.3%, 97.3%)
mixture model – IU/ml	85.9%	91.0%	93.3%	95.6%
	(82.8%, 88.5%)	(88.6%, 93.0%)	(91.8%, 94.5%)	(93.3%, 97.2%)

* 95% Binomial exact CI

9.10 Mixture model on OD measurements adjusting for plate-to-plate variability

An alternative method to adjust for plate-to-plate variability is to fit a mixture model whose parameters vary by plate. Extending Equation 7.16 to account for plate-to-plate variation - assuming random allocation of samples to each plate - the mixture model becomes of the form

$$g_{ii}(x) = (1 - \pi_i) f_N(x; \mu_i, \sigma) + \pi_i f_{SN}(x; v_{ii}, \tau_i, \lambda_q),$$

Equation 9.5

where $i = 1, \dots, \zeta$ are the age groups and $t = 1, \dots, \psi$ are the plates. Therefore, in addition to varying the parameters of the positive component by age group, the location parameters of both components vary by plate.

The problem with this method is that it requires large numbers of parameters. For the UK rubella serosurvey example, there are 4 age groups and 44 plates (after the exclusion of plate 38). This translates to 5 additional parameters for each plate i.e. 233 parameters would be needed (44 parameters for μ_t , 1 for σ , $44 \times 4 = 176$ for v_{it} , 4 for τ_t , 4 for λ_t and 4 parameters for π_i).

A solution to this problem is to restrict the parameters, by grouping the plates using the OD measurement of the control samples on each plate. For the UK rubella example the 44 plates were divided into 6 groups as shown in Table 9.4.

Table 9.4 Plate groups with respect to negative controls

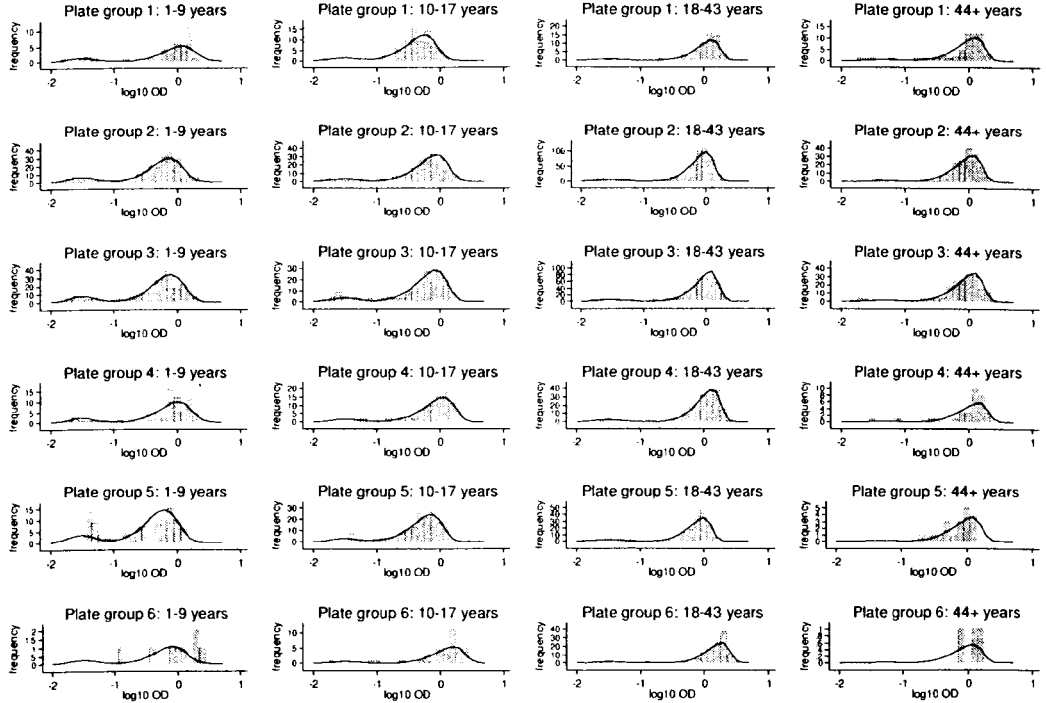
OD of (negative)		
control sample	plates	group
0.009 - 0.011	9, 10, 29	1
0.012 - 0.014	1, 3, 5, 14, 20, 21, 22, 25, 26, 27, 30, 32, 41, 45	2
0.015 - 0.017	4, 7, 8, 12, 13, 16, 17, 19, 28, 31, 35, 36, 42, 44	3
0.018 - 0.023	2, 33, 34, 37, 43	4
0.027 - 0.036	15, 18, 23, 24, 39, 40	5
0.05 - 0.075	6, 11	6

Using the mixture model expressed in Equation 9.5 for $\zeta = 4$ and $\psi = 6$, the resulting estimates were $\hat{\mu}_1 = -1.55$, $\hat{\mu}_2 = -1.54$, $\hat{\mu}_3 = -1.56$, $\hat{\mu}_4 = -1.40$, $\hat{\mu}_5 = -1.37$, $\hat{\mu}_6 = -1.32$, $\hat{\sigma} = 0.21$ for the negative component, and $\hat{v}_{11} = 0.23$, $\hat{v}_{21} = -0.10$, $\hat{v}_{31} = 0.21$, $\hat{v}_{41} = 0.21$, $\hat{v}_{12} = 0.02$, $\hat{v}_{22} = 0.08$, $\hat{v}_{32} = 0.09$, $\hat{v}_{42} = 0.17$, $\hat{v}_{13} = 0.05$, $\hat{v}_{23} = 0.06$, $\hat{v}_{33} = 0.16$, $\hat{v}_{43} = 0.19$,

$\hat{v}_{14} = 0.14$, $\hat{v}_{24} = 0.17$, $\hat{v}_{34} = 0.21$, $\hat{v}_{44} = 0.26$, $\hat{v}_{15} = -0.05$, $\hat{v}_{25} = -0.01$, $\hat{v}_{35} = 0.08$, $\hat{v}_{45} = 0.18$,
 $\hat{v}_{16} = 0.07$, $\hat{v}_{26} = 0.34$, $\hat{v}_{36} = 0.35$, $\hat{v}_{46} = 0.18$, $\hat{\tau}_1 = 0.39$, $\hat{\tau}_2 = 0.40$, $\hat{\tau}_3 = 0.31$, $\hat{\tau}_4 = 0.37$,
 $\hat{\lambda}_1 = -2.43$, $\hat{\lambda}_2 = -3.28$, $\hat{\lambda}_3 = -4.17$, $\hat{\lambda}_4 = -4.43$ for the positive component. Note that there
 were not enough samples for ages greater than 44 years in plate groups 5 and 6, and
 similarly, for age group 1-9 in plate group 6. In these cases fixed parameters were used
 based on the simpler model (Equation 7.16).

Figure 9.14 shows the observed frequencies and fitted distribution by age and plate groups. The low number of observations, which influence the model fit is depicted clearly for plate groups 5 and 6.

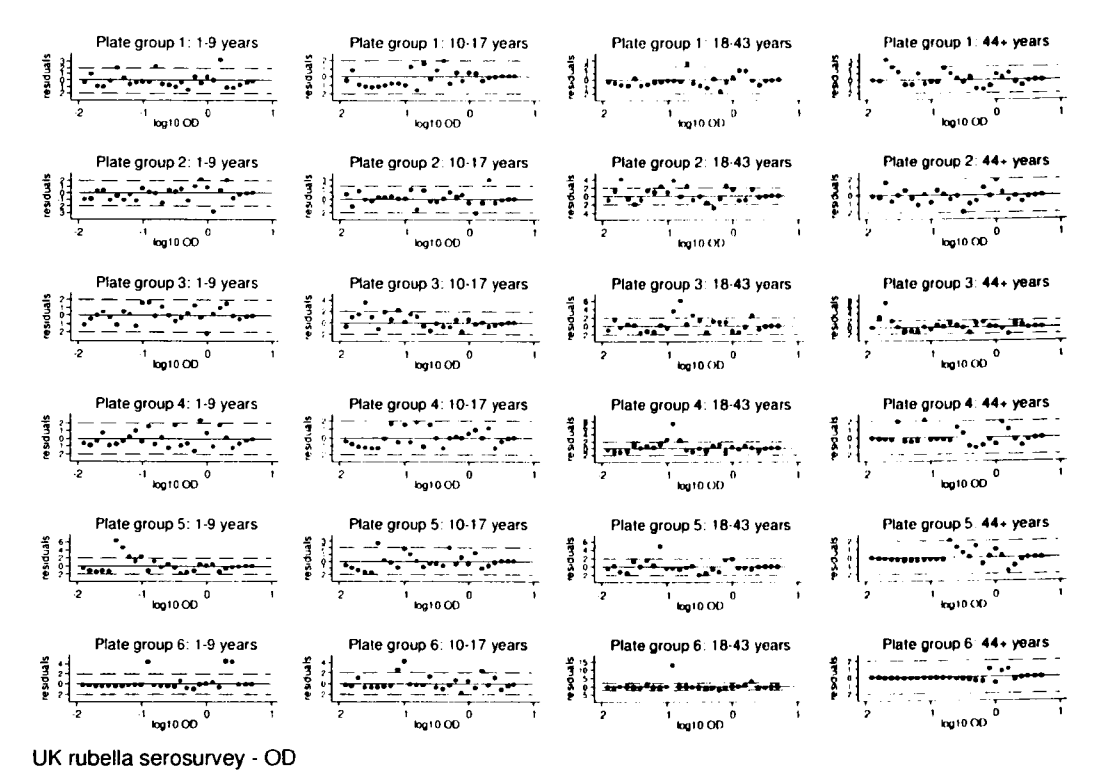
Figure 9.14 Observed frequencies and fitted mixture model by age and plate groups



UK rubella serosurvey - OD

Figure 9.15 shows the Pearson residuals plotted against the \log_{10} OD measurements by age and plate groups. The residual plots indicate a lack of fit in the area -1 to -0.8 (on the \log_{10} OD scale), where the underlying distribution components meet for ages 18 to 43 years, especially for plate groups 3, 4, 5 and 6. Residuals far outside the ± 2 limits were also apparent for the negative underlying components for plate 3 and those older than 44 years of age and also for plate 5 and ages between 1 and 9 years. In these cases fixed parameters were used.

Figure 9.15 Pearson residuals by age group and plate groups for \log_{10} OD



The seroprevalence estimated from Equation 9.5 was 84.2%, 90.9%, 93.3% and 95.7% for age groups 1-9, 10-17, 18-43 and 44+, respectively. These estimates are closer to the seroprevalences estimated using the IU/ml, which suggests that this may be a better method to adjust for plate-to-plate variability than the adjusted OD.

9.11 Conclusions

The aim of this Chapter was to investigate the effect that plate-to-plate variability has on seroprevalence estimates. A mixture model fitted on the crude OD measurements (not adjusted for plate-to-plate variability) gave slightly different seroprevalence estimates than the IU/ml where any bias had been corrected by calibration curves. The adjusted OD measurements (OD measurements divided by a control sample in each plate) gave similar estimates to the crude OD, which suggests that this method might not be sufficient to correct for plate-to-plate variability.

A new method was also proposed which involves adjusting for plate-to-plate variability by fitting separate mixture models to the crude OD measurements on each plate. The main disadvantage of this method is that it requires more complex models which may lead to a high number of parameters, depending on how many samples are tested on each plate (this problem may be overcome by grouping the plates as shown in Section 9.10).

Another method which was not examined would be to treat the plates as random effects assuming two sources of variability, i.e. between-measurements and between-plates. Then seroprevalence can be estimated by adjusting plate-to-plate variability using a *random-effect* mixture model.

Adjusting for plate-to-plate variability during mixture modelling could be a useful method for estimating seroprevalence. By avoiding testing for control samples one could reduce the cost of a serosurvey. For example, for the UK serosurvey tested against rubella, a serological plate contained 96 wells, 8 of which were filled by control samples for calculating the calibration equations. By not testing these samples and using the crude OD results, it would require testing 8.3% fewer samples which is an important saving for serosurveys containing a large numbers of samples. However, the idea behind this method

that is based on mixture models is to be used as a complementary method to the existing method using calibration curves. As mentioned in Chapter 5, mixture modelling is based on certain assumptions that, when violated, may result in biased estimates.

In the UK example presented, small differences (less than 2%) in seroprevalence estimates were found using different mixture modelling methods. These may not be of major concern from a public health point of view and suggest that there is little effect of plate-to-plate variability on seroprevalence estimation. However, it is important to stress that these conclusions are based on one serosurvey where a specific rubella assay was used. It is likely that the impact of plate-to-plate variability on seroprevalence estimates will be higher for other assays or antigens. Therefore, further research on the issue is necessary before more definite conclusions can be drawn.

Chapter 10: Discussion

10.1 Introduction

Infectious diseases account for 23.4% of the global deaths, and this proportion is continuously decreasing (using 2004 mortality estimates as baseline) (World Health Organization, 2008d). Much of the mortality reduction can be attributed to the success of vaccines, which could have been much greater if more efficient vaccination programmes had been applied (World Health Organization, 2008d; World Health Organization, 2009). Even developed countries with long-established vaccination campaigns have morbidity and mortality resulting from vaccine-preventable diseases. WHO estimates that, in Europe, approximately 32,000 deaths could have been prevented by more effective use of existing vaccines (World Health Organization, 2009).

The ESEN2 project was a European study funded by the European Commission in 2001. The project aimed to identify susceptible cohorts for a number of vaccine-preventable infections across Europe by harmonizing serological markers. The main benefit of such harmonisation for each participant country is that this information may be used to target and subsequently amend any weaknesses in vaccination campaigns. In addition, the comparison of the various vaccination programmes between different countries may lead to optimisation of vaccination strategies at an international level.

In this thesis, the information collected as part of the ESEN2 study was used to establish a statistical-based algorithm for standardising serological outcomes tested in different laboratories into comparable units. This algorithm - based on methods used in the past - was employed during the ESEN2 project. Initially the impact of data outside the assay detection limits was examined and, after comparing different statistical techniques, censored regression was proposed for improving standardisation equations. The impact of the standardisation algorithm on seroprevalence estimation was examined. Differences

between standardised and non-standardised seroprevalence were mainly attributed to differences between antigens and laboratory procedures.

Mixture modelling was proposed as an alternative method to standardisation for estimating population seroprevalence. Although mixture modelling may provide better seroprevalence estimates in certain situations, it is heavily dependent on specific assumptions, mainly of well-separated underlying component distributions. In terms of seroprevalence estimation using standardisation, the validity of the assay cut-off point was examined. A method for estimating cut-offs was proposed based on mixture modelling. This method might improve seroprevalence assuming the distributional assumptions hold.

The impact of plate-to-plate variability, occurring due to samples tested in batches, on seroprevalence estimates was also examined. A comparison of seroprevalence estimates based on the raw OD results and the plate-adjusted IU measurements showed very small seroprevalence differences. An alternative method to adjust for plate-to-plate variability without transforming OD results into IU was proposed, again through the use of mixture modelling.

10.2 Standardisation methodology

The main aim of the first part of the thesis was to establish a statistical algorithm for obtaining regression equations. In the ESEN2 project these would be used for standardising serological results into common units based on the statistical methods employed in a previous European study (ESEN). Other objectives were to validate and further develop the algorithm and evaluate the impact of standardised units on seroprevalence estimation.

Twenty two countries participated in the ESEN2 project by collecting and testing national serosurveys of between 1,000 and 3,500 blood samples for all or some of the eight infections included in the project. In order to standardise the serological output, standardisation panels of approximately 150 samples - ranging from negative and equivocal (or low positive) to positive samples - were tested by a reference laboratory for each antigen. These panels were subsequently sent and tested by each participant laboratory using the assay method of their choice (Andrews N *et al.*, 2000). By regressing national panel test results against those of the reference centre, equations were generated which were used for standardising the main serosurvey results.

There were three statistical issues affecting standardisation equation estimates, namely (a) extreme measurements (outliers), (b) type of the regression model and (c) effect of measurements outside the quantitative assay limits (censored observations).

A statistical algorithm was developed to deal with these issues in three steps: (a) a quadratic regression equation was initially fitted and any observations with inflated standardised residuals were flagged as outliers. Following an investigation to ensure that no error existed during testing or data entry, these measurements were omitted from the analysis in a two-step process. (b) Linear and quadratic equations were considered. The simple linear regression model was generally used unless there was significant evidence that the quadratic term provided a better fit. In the few cases, where a clear lack of fit by both models was observed, a sigmoid-type model was used. (c) Two methods were considered for taking into account measurements outside the assay detection limits: The selected regression model was fitted once after omitting censored observations and again after substituting these with a constant ($D_L / 2$ and $2D_U$ for left- and right-censored data, respectively). In most cases the simple substitution method was used. However, in situations where substituted measurements greatly influenced the model in the area of

interest where misclassification was likely to occur (positive / negative cut-off for the reference assay), censored observations were omitted from the regression model.

The statistical algorithm described above was developed as part of this thesis and was used to standardise serological outcomes for the ESEN2 project. The validity of this algorithm, in terms of censored outcomes, was examined in detail and alternative methods were proposed. Apart from the methods of deletion and simple substitution employed for the ESEN2 project, censored regression, multiple imputation and other simple substitution methods were considered. Simulation results showed that although all methods used in the ESEN2 project gave estimates close to truth, and therefore may be considered valid, censored regression and multiple imputation methods provided still more robust estimates, especially for higher amounts of censored data. Moreover, interval censored regression was suggested for serological measurements in the form of dilution series. Finally, a non-parametric test was proposed for confirming the validity of a regression equation for x-censored data.

Seroprevalence estimates were obtained by classifying serological results as positive or negative according to a cut-off point. The impact of using standardised units on seroprevalence estimates was evaluated by comparing the estimates before and after standardisation was implemented. The median seroprevalence change following standardisation was 1.6 (IQR: [0.4, 5.1]) and the geometric mean 3.5 (95% CI: [2.5, 4.5]). The results showed differences in seroprevalence estimates, especially between antigens, with mumps and HAV having the highest numbers of samples re-classified following standardisation. There were important changes between seroprevalence estimates following standardisation between laboratories.

10.3 Seroprevalence estimation

The second part of the thesis aimed at validating the techniques used during the ESEN2 project for obtaining seroprevalence estimates and to propose alternative methods. A further objective was to investigate the impact of plate-to-plate variability on seroprevalence estimation.

Seroprevalence estimation as used in the ESEN2 project depended on two assumptions: (a) as discussed in Section 10.2, successful standardisation of the serological data and (b) a suitable choice of the assay cut-off. Both these assumptions may be bypassed when using an alternative method for seroprevalence estimation i.e. mixture modelling.

Using mixture modelling directly on non-standardised serological data makes it possible to obtain seroprevalence estimates without applying assay cut-offs, and therefore, avoiding the time-consuming and costly process of standardisation. However, there are two major disadvantages associated with this method: (a) there is a certain element of subjectivity associated with choosing the shape (and the number) of the underlying components of mixture models and (b) the distributions of the underlying components need to be well-separated which was not always the case for the serosurveys examined in the project.

For the ESEN2 project, the reference centre's assay cut-off was used to classify the standardised units into positive and negative. The suitability of the assay manufacturer's cut-offs for seroepidemiology purposes have been questioned in the past (Vyse AJ *et al.*, 2004). Therefore, an iterative method for estimating age-specific assay cut-offs for seroprevalence estimation was proposed. The cut-off was defined as the point that generated equal numbers of false positive and false negative samples, based on the underlying components of a mixture model.

Plate-to-plate variability was taken into account during the ESEN2 serosurvey tests by including control samples of known IU measurements on each plate. Calibration curves were generated from these control samples after regressing the known IU measurements against the ODs. The calibration curves were subsequently used to transform the OD measurements in each plate into IU/ml. There was little impact of this variability adjustment on seroprevalence estimation on the example examined. A novel method was proposed to adjust for plate-to-plate variability using mixture modelling without having to generate calibration curves.

10.4 Limitations of the standardisation methodology

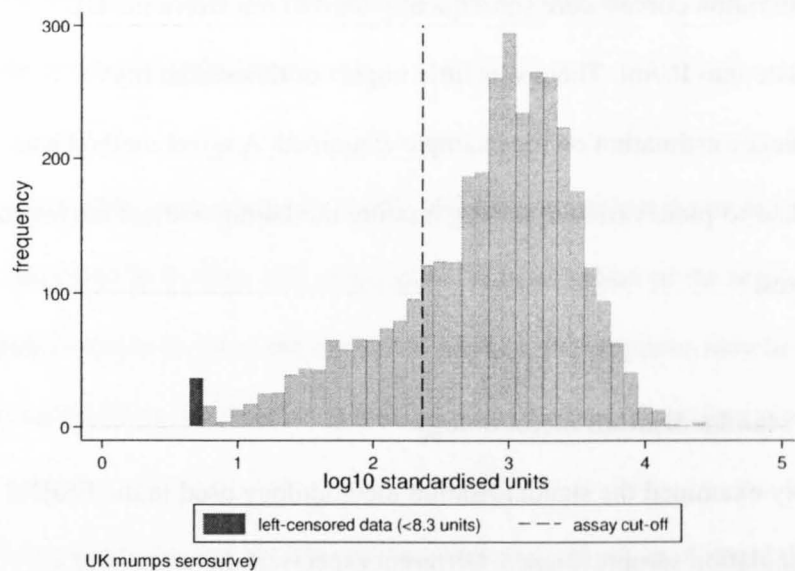
This thesis has thoroughly examined the standardisation methodology used in the ESEN2 project for estimating population seroprevalence. Different aspects of the process of standardising serological results were investigated and validated. Obtaining seroprevalence estimates through the application of the assay manufacturer's cut-off point was assessed.

Examples of distributions with well-separated underlying components presented in this thesis show that a manufacturer's cut-off is usually suitable for seroprevalence estimation, and is often close to the optimal cut-off. If the cut-off is slightly biased, the effect on seroprevalence estimates may be limited, if there is small number of samples around the area where the underlying components meet.

However, in situations where the underlying distributions are not well separated - such as the mumps serosurveys - the appropriateness of the assay cut-off for seroepidemiology is not so obvious. In the example introduced in Section 7.15, showing the distribution of the UK serosurvey samples tested for mumps, the underlying components were not well separated. The titre distribution together with the assay cut-off point is given in Figure

10.1. Given the large number of samples around the critical area where the cut-off lies, it is clear that a bias in the cut-off will greatly affect the seroprevalence estimates.

Figure 10.1 Distribution of the mumps serosurvey results for UK and assay cut-off



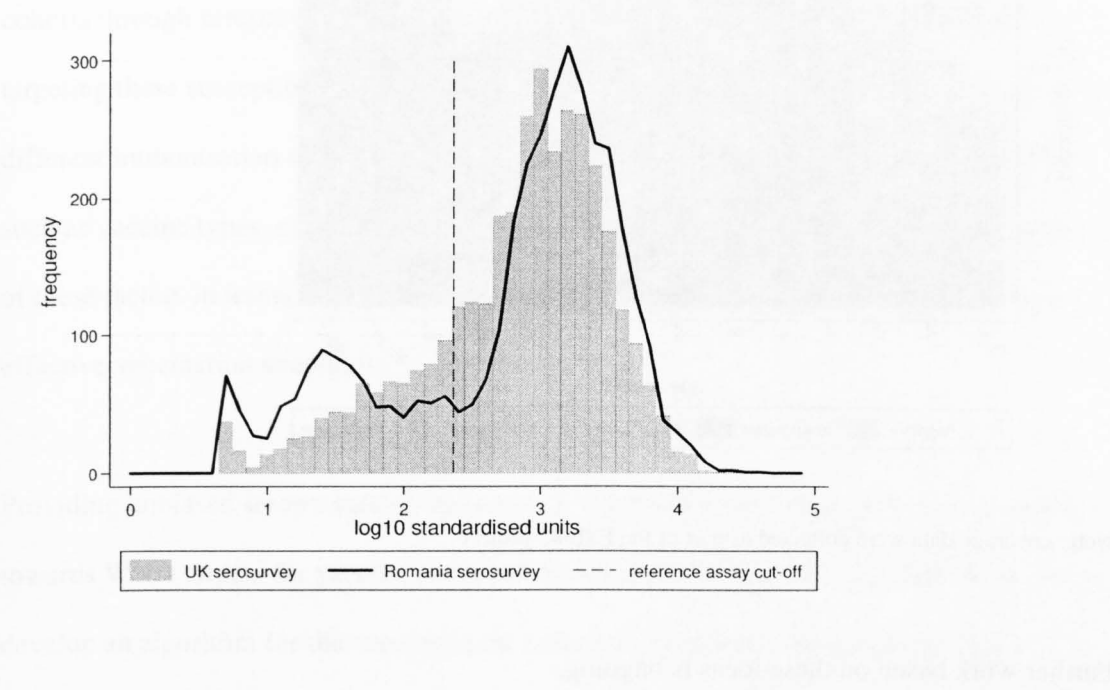
10.5 Extending mixture modelling methodology: current work

The reason why there is no clear separation between the negative and positive underlying components could be due to the presence of vaccinated individuals. In theory, these could be represented by a third, “intermediate”, component, and separate estimates for the proportion of individuals vaccinated, past infected and non-infected / non-vaccinated may be obtained by mixture modelling. However, this is not a practical option for the UK results, since there is no separation between the distributions of the underlying components.

Vaccinated groups are present in all national serosurveys included in ESEN2, apart from Romania, where there was no routine vaccination against mumps when the serosurvey was collected (2002). Therefore, the distribution of Romanian results may serve to define the underlying distributions of the UK serosurvey, by helping to identify an additional

component for the vaccinated group. Figure 10.2 shows the distribution of the standardised Romanian and UK mumps results. The samples from the UK serosurvey measuring between 1.8 and 2.7 on the \log_{10} -scale may correspond to weaker (compared to natural infection) immunological responses induced by vaccination.

Figure 10.2 Distribution of the mumps serosurvey results for UK and Romania and reference centre’s assay cut-off

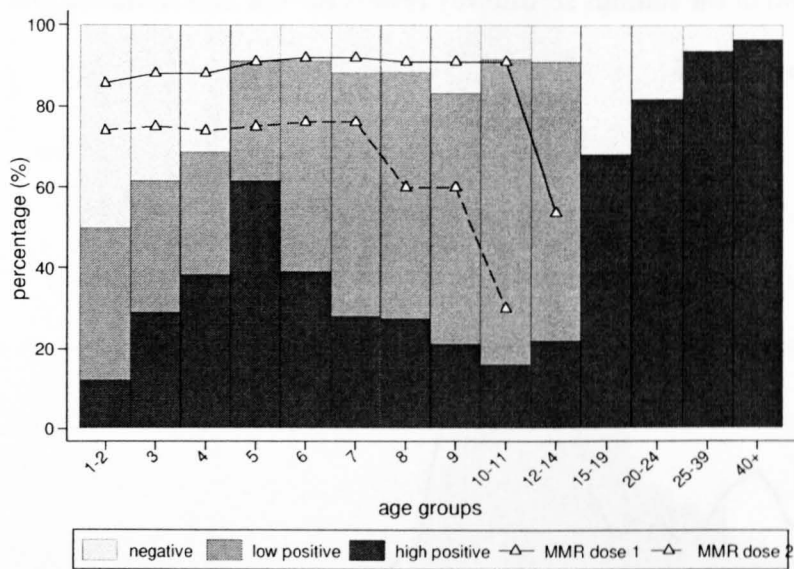


This suggests an extension of the mixture modelling presented in this thesis, using data from countries with contrasting vaccination schedules to identify the components of mixture models.

Figure 10.3 shows the results from an initial analysis of the UK mumps seroprevalence data along these lines using mixture modelling. Seroprevalence by age is presented in combination with vaccine coverage of the first and second MMR dose. This suggests that high vaccine coverage corresponds to lower proportions of natural infections (high positive and higher proportions of positives). For ages older than 15 years, high titre measurements

may be attributed to natural infection, given that no vaccination programme had been introduced at the time.

Figure 10.3 Estimates of negative, vaccinated and past infected groups as well as MMR vaccine coverage by age group



Note: coverage data were collected as part of the ESEN2 project

Further work based on these ideas is ongoing.

10.6 Future applications

This research demonstrates the validity of the methods used for the ESEN2 project and proposes some improvements, as well as new methods for standardising serological outcomes and estimating population seroprevalence. An obvious application of the techniques evaluated and developed during this thesis would be a future seroepidemiology project. Initial discussions include proposals for a new project that could include a number of additional infections. Examples of infections could be Hepatitis C Virus (HCV), Human Papillomavirus (HPV) or Haemophilus Influenzae type B (HIB).

Apart from a future European study similar to the ESEN and ESEN2 projects, the methods developed during this thesis could be used in a variety of seroepidemiological applications, mainly with regards to assay comparison and seroprevalence estimation.

10.7 General conclusions

Seroepidemiological studies play an important role in identifying susceptible population cohorts through seroprevalence estimation, and in improving vaccination campaigns by targeting these susceptible groups. International seroepidemiological studies comparing different immunisation strategies may provide valuable information for assessing factors such as vaccine types, number of doses or age at vaccination. Quantification of the impact of these factors in terms of population seroprevalence can aid the development of more effective vaccination strategies.

Providing unbiased seroprevalence estimates is very important for assessing the progress towards WHO targets for vaccine-preventable diseases in Europe. Part of this thesis was to develop an algorithm for the standardisation method. Standardisation was used for obtaining seroprevalence estimates as part of the ESEN2 project and was validated by the work presented here. Moreover, different aspects of this method were examined and advanced in order to be used in future studies. A future addition to the standardisation algorithm might include the use of censored regression models. The impact of the algorithm on seroprevalence estimation was highlighted by estimating seroprevalence prior to and following standardisation.

Other methods for obtaining seroprevalence estimates such as mixture modelling were also considered. In many cases, mixture modelling may provide better estimates *compared to* the standardisation method that is based on fixed assay cut-offs. However, in order to

obtain robust estimates using mixture modelling, clear distinction between the underlying distributions is needed. Unfortunately this is not the case when sampling from a vaccinated population, where vaccine-induced antibody response is weaker than natural infection and wanes in time. In general, although the fixed cut-offs used during the ESEN2 project can be considered valid, a method was introduced for estimating age-specific cut-offs in an iterative way based on the reference centre's serosurvey distribution. However, this method is heavily dependent on the assumption of similar location and dispersion parameters for the underlying distributions between the reference and the participant laboratories.

Finally, the impact of the variability occurring by testing the serological samples in batches (plate-to-plate variability) on seroprevalence was assessed. Using a VZV assay example, there was maximum 2% difference in seroprevalence after using the variability-adjusted IU/ml measurements. A new method was proposed to account for plate-to-plate variability in seroprevalence estimation without having to estimate calibration curves.

In conclusion, the standardisation method as used in the ESEN2 project is a simple and robust method for obtaining seroprevalence when combining samples tested in different laboratories by various assays. Therefore, following some method improvements suggested in this thesis, standardisation would be the preferred option to use for similar projects in the future.

Reference List

Ades AE (1990). Evaluating screening tests and screening programmes. *Archives of Disease in Childhood* **65**, 792-795.

Akritis MG, Murphy SA, & LaValley Mp (1995). The Theil-Sen estimator with doubly censored data and applications to astronomy. *Journal of the American Statistical Association* **90**, 170-177.

Anastassopoulou CG, Kafatos G, Nardone A, Andrews N, Pebody R, Mossong J, Davidkin I, Gelb D, de Ory F, Thierfelder W, Nemecek V, Bruzzzone B, Butur D, Barbara C, Sobotova Z, Jones L, Griskevicius A, Hesketh LM, Cohen D, Vranckx R, Tsakris A, Miller E. & Hatzakis A (2009). The European Sero-Epidemiology Network 2 (ESEN2): standardisation of assay results for hepatitis A virus (HAV) to enable comparisons of seroprevalence data across 15 countries. *Epidemiology and Infection* **137**, 485-494.

Anderson RM & May RM (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford.

Andrews N, Pebody RG, Berbers G, Blondeau C, Crovari P, Davidkin I, Farrington P, Fievet-Groyne F, Gabutti G, Gerike E, Giordano C, Hesketh L, Marzec T, Morgan-Capner P, Osborne K, Pleisner AM, Raux M, Tischler A, Ruden U, Valle M, & Miller E (2000). The European Sero-Epidemiology Network: standardising the enzyme immunoassay results for measles, mumps and rubella. *Epidemiology and Infection* **125**, 127-141.

Andrews N, Stowe J, Wise L, & Miller E (2010). Post licensure comparison of the safety profile of diphtheria / tetanus / whole cell pertussis / haemophilus influenza type b vaccine and a 5-in-1 diphtheria / tetanus / acellular pertussis / haemophilus influenza type b / polio vaccine in the United Kingdom. *Vaccine* **28**, 7215-7220.

Andrews N, Tischer A, Siedler A, Pebody RG, Barbara C, Cotter S, Duks A, Gacheva N, Bohumir K, Johansen K, Mossong J, de Ory F, Prosenc K, Slacikova M, Theeten H, Zarvou M, Pistol A, Bartha K, Cohen D, Backhouse J, Griskevicius A, & Nardone A (2008). Towards elimination: measles susceptibility in Australia and 17 European countries. *Bulletin of World Health Organization* **86**, 197-204.

Armitage P & Berry G (1994). *Statistical Methods in Medical Research*. Blackwell Science Ltd, Oxford.

Azzalini A (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171-178.

Baccarelli A, Pfeiffer R, Consonni D, Pesatori AC, Bonzini M, Patterson DG, Bertazzi PA, & Landi MT (2005). Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chlorance study. *Chemosphere* **60**, 898-906.

Barnhart HX, Song J, & Lyles RH (2005). Assay validation for left-censored data. *Statistics in Medicine* **24**, 3347-3360.

Baughman AL, Bisgard KM, Lynn F, & Meade BD (2006). Mixture model analysis for establishing a diagnostic cut-off point for pertussis antibody levels. *Statistics in Medicine* **25**, 2994-3010.

Biernacki C, Celeux G, & Govaert G (2000). Assessing a mixture model for clustering with integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 719-725.

Bland JM & Altman DG (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **327**, 307-310.

Bland M (2000). *An Introduction to Medical Statistics*. Oxford University Press, Oxford.

Breen R (1996). *Regression Models: Censored, Sample Selected, or Truncated Data*. Sage publications, Thousand Oaks, California.

Brent R (1973). *Algorithms for Minimisation Without Derivatives*. Prentice-Hall, Englewood Cliffs, NJ.

Brisson M & Edmunds WJ (2003). Varicella vaccination in England and Wales: cost-utility analysis. *Archives of Disease in Childhood* **88**, 862-869.

Brockhoff HJ, Mollema L, Sonder GJB, Postema CA, van Binnendijk RS, Kohl RHG, de Melker HE, & Hahné SJM (2010). Mumps outbreak in a highly vaccinated student population, The Netherlands, 2004. *Vaccine* **28**, 2932-2936.

Carpenter J & Goldstein H (2004). Multiple imputation in MLwiN. *Multilevel Modelling Newsletter* **16**, 9-18.

Centers for Disease Control and Prevention (2010). Update: mumps outbreak - New York and New Jersey. June 2009-January 2010. *Morbidity and Mortality Weekly Report* **59**, 125-129.

Centre for Disease Control and Prevention (2004). Progress toward elimination of measles and prevention of congenital rubella infection - European Region, 1990-2004. *Morbidity and Mortality Weekly Report* **54**, 175-178.

Chen RT, Markowitz LE, Albrecht P, Stewart JA, Mofenson LM, Preblud SR, & Orenstein WA (1990). Measles antibody: Re-evaluation of protective titres. *Journal of Infectious Diseases* **162**, 1036-1042.

Cianciara J (2000). Hepatitis A shifting epidemiology in Poland and Eastern Europe. *Vaccine* **18**, S70.

Collett D (1999). *Modelling Binary Data*. Chapman & Hall / CRC, Boca Raton.

Daniels RD & Yiin JH (2006). A comparison of statistical methods for estimation of less than detectable ionising radiation exposures. *Radiation Protection Dosimetry* **121**, 240-251.

Davies DH, Halablab MA, Clarke J, Cox FEG, & Young TWK (1999). *Infection and Immunity*. Taylor and Francis, London.

de Melker H, Conyn-van Spaendonck M, Rumke HC, van Wijngaarden JK, Mooi FR, & Schellekens JFP (1997). Pertussis in the Netherlands: an outbreak despite high levels of immunisation with whole-cell vaccine. *Emerging Infectious Diseases* **3**, 175-178.

de Melker H, Pebody RG, Edmunds WJ, Lévy-Bruhl D, Valle M, Rota MC, Salmaso S, van de Hof S, Berbers G, Saliou P, Conyn-van Spaendonck M, Crovari P, Davidkin I, Gabutti G, Hesketh L, Morgan-Capner P, Plesner AM, Raux M, Tischer A, & Miller E

(2001). The seroepidemiology of measles in Western Europe. *Epidemiology and Infection* **2001**, 249-259.

de Ory F, Echevarria JM, Kafatos G, Anastassopoulou C, Andrews N, Backhouse J, Berbers G, Bruckova B, Cohen DI, de Melker H, Davidkin I, Gabutti G, Hesketh LM, Johansen K, Jokinen S, Jones L, Linde A, Miller E, Mossong J, Nardone A, Rota MC, Sauerbrei A, Schneider F, Smetana Z, Tischer A, Tsakris A, & Vranckx R (2006). European seroepidemiology network 2: Standardisation of assays for seroepidemiology of varicella zoster virus. *Journal of Clinical Virology* **36**, 111-118.

Dong Z (1997). Mixture analysis and its preliminary application in archaeology. *Journal of Archaeological Science* **24**, 141-161.

Dorland (2000). *Dorland's Medical Dictionary*., 29th ed. W.B. Saunders company, Philadelphia.

Dowdle WR (1999). The principles of disease elimination and eradication. *Morbidity and Mortality Weekly Report* **48**, 23-27.

Edmunds WJ, Pebody RG, Aggerbeck H, Baron S, Berbers G, Conyn-van Spaendonck M, Hallander HO, Olander RM, Maple PAC, de Melker H, Olin P, Fievet-Groyne F, Rota MC, Salmaso S, Tischer A, von Hunolstein C, & Miller E (2000). The sero-epidemiology of diphtheria in Western Europe. *Epidemiology and Infection* **125**, 113-125.

Edmunds WJ, Pebody RG, Aggerbeck H, Baron S, Berbers G, Conyn-van Spaendonck M, Hallander HO, Olander RM, Maple PAC, de Melker H, Olin P, Fievet-Groyne F, Rota MC, Salmaso S, Tischer A, von Hunolstein C, & Miller E (2001). Erratum. *Epidemiology and Infection* **126**, 331.

Efron B & Tibshirani RJ (1993). *An Introduction to the Bootstrap*. Chapman and Hall, London.

Eurosurveillance editorial team (2007). Measles deaths fall by 60 percent worldwide. *Eurosurveillance* **12**, article=1.

Farrington CP (1990). Modelling forces of infection for measles, mumps and rubella. *Statistics in Medicine* **9**, 953-967.

Fox J (1991). *Regression Diagnostics*. Sage publications, California.

Galazka A & Tomaszunas-Blaszczyk J (1997). Why do adults contract diphtheria? *Eurosurveillance* **2**, article=1.

Galazka AM, Robertson SE, & Oblapenko GP (1995). Resurgence of diphtheria. *European Journal of Epidemiology* **11**, 95-105.

Gay NJ (1996). Analysis of serological surveys using mixture models: application to a survey of parvovirus B19. *Statistics in Medicine* **15**, 1567-1573.

Giammanco A, Chiarini A, Maple PAC, Andrews N, Pebody R, Gay N, Olander RM, Fivet-Groyne F, Baron S, Tischer A, Swidsinski S, Schellekens J, & Reizenstein E (2003). European Sero-Epidemiology Network: standardisation of the assay results for pertussis. *Vaccine* **22**, 112-120.

Giammanco A, Nardone A, Pebody R, Kafatos G, Andrews N, Chiarini A, Taormina S, de Ory F, Prosenc K, Kriz B, Hallander H, Ljungman M, Marva E, Tsakris A, O'Flanagan D, Schneider F, Griskevicius A, Vranckx R, & Karacs I (2008). European Sero-Epidemiology

Network 2: Standardisation of immunoassay results for pertussis requires homogeneity in the antigenic preparations. *Vaccine* **26**, 4486-4493.

Giesecke J (2002). *Modern Infectious Disease Epidemiology*, 2nd ed. Arnold, London.

Gould W & Sribney W (1999). *Maximum Likelihood Estimation with Stata*. Stata Press, Texas.

Greene WH (2003). Limited dependent variable and duration models. In *Econometric Analysis* pp. 756-790. Prentice Hall, New Jersey.

Hallander HO & Gustafsson L (2009). Efficacy and effectiveness of ecellular pertussis vaccines: a 20-year Swedish experience. *Expert Review of Vaccines* **8**, 1303-1307.

Hardelid P (2008). *The Seroepidemiology of Rubella in Pregnant Women in North Thames* Thesis (PhD) University of London.

Hardy IR, Dittman S, & Sutter RW (1996). Current situation and control strategies for resurgence of diphtheria in Newly Independent States of the former Soviet Union. *Lancet* **347**, 1739-1744.

Helsel DR (2005). More than obvious: better methods for interpreting nondetect data. *Environmental Science and Technology* **39**, 419A-423A.

Holden D (2004). testing the normality assumption in the Tobit model. *Journal of Applied Statistics* **31**, 521-532.

- Jacobsen KH & Koopman JS (2004). Declining hepatitis A seroprevalence: a global review and analysis. *Epidemiology and Infection* **132**, 1005-1022.
- Jain RB, Caudill SP, Wang RY, & Monsell E (2008). Evaluation of maximum likelihood procedures to estimate left censored observations. *Analytical Chemistry* **80**, 1124-1132.
- Kafatos G, Andrews N, & Nardone A (2005). Model selection methodology for inter-laboratory standardisation of antibody titres. *Vaccine* **23**, 5022-5027.
- Kirkwood BR & Sterne JAC (2003). *Essential Medical Statistics*, 2nd ed. Blackwell, Malden Massachusetts.
- Kolodkina V, Titov L, Sharapa T, Grimont F, Grimont PAD, & Efstratiou A (2006). Molecular epidemiology of *C. diphtheriae* strains during different phases of diphtheria epidemic in Belarus. *MBC Infectious Diseases* **6**.
- Krishnamoorthy K, Mallick A, & Mathew T (2009). Model-based imputation approach for data analysis in the presence of non-detects. *Annals of Occupational Hygiene* **53**, 249-263.
- Kuha J (2004). AIC and BIC: Comparisons of assumptions and performance. *Sociological Methods & Research* **33**, 188-229.
- Lambert PH, Liu M, & Siegrist CA (2005). Can successful vaccines teach us how to induce efficient protective immune responses? *Nature medicine* **11**, S54-S62.
- Lim J (2006). Permutation procedures with censored data. *Computational Statistics and Data Analysis* **50**, 332-345.

Lubin JH, Colt JS, Camann D, Davis S, Cerhan JR, & Severson RK (2004).

Epidemiological evaluation of measurement data in the presence of detection limits.

Environmental Health Perspectives **112**, 1691-1696.

Lyles RH, Williams JK, & Chuachoowong R (2001). Correlating two viral load assays with known detection limits. *Biometrics* **57**, 1238-1244.

Lynn HS (2001). Maximum likelihood inference for left-censored HIV RNA data.

Statistics in Medicine **20**, 33-45.

Maddala GS (2001). *Introduction to Econometrics*, 3rd edition ed., pp. 333-338. John Wiley and Sons, Chichester.

Male D, Brostoff J, Roth DB, & Roitt I (2006). *Immunology*, 7th ed ed. Mosby Elsevier, Canada.

Maple PAC, Simms I, Kafatos G, Solomou M, & Fenton K (2006). Application of a noninvasive oral fluid test for detection of treponemal IgG in a predominantly HIV-infected population. *European Journal of Clinical Microbiology and Infectious Diseases* **25**, 743-749.

Martin R, Deshevoi S, Buddha N, & Jankovic D (2009). Approaching measles and rubella elimination in the European Region - need to sustain the gains. *Eurosurveillance* **14**, article=1.

McLachlan G & Peel D (2000). *Finite Mixture Models*. John Wiley & Sons, Inc, new York.

Meerhoff T, Meijer A, & Paget W (2004). Methods for sentinel virological surveillance of influenza in Europe - an 18-country survey. *Eurosurveillance* **9**, article=4.

Miller E (1991). Rubella in the United Kingdom. *Epidemiology and Infection* **107**, 31-42.

Mims C, Playfair J, Roitt I, Wakelin D, & Williams R (1998). *Medical Microbiology*. Mosby, London.

Morgan BJT (2000). *Applied Stochastic Modelling*. Arnold, London.

Mosby. Mosby's medical dictionary. <http://medical-dictionary.thefreedictionary.com/prevalence> . 2009. Elsevier. Access date. 7-9-2010.

Ref Type: Electronic Citation

Mossong J, Hens N, Friederichs V, Davidkin I, Broman M, Litwinska B, Siennicka J, Trzcinska A, Van Damme P, Beutels P, Vyse A, Shkedy Z, Aerts M, Massari M, & Gabutti G (2007). Parvovirus B19 infection in five European countries: seroepidemiology, force of infection and maternal risk of infection. *Epidemiology and Infection* **136**, 1059-1068.

Muscat M, Bang H, Wohlfahrt J, Glismann S, & Molbak K (2009). Measles in Europe: an epidemiological assessment. *Lancet* **373**, 383-389.

Nardone A, Anastassopoulou CG, Theeten H, Kriz B, Davidkin I, Thierfelder W, O'Flanagan D, Bruzzone B, Mossong J, Boot HJ, Butur D, Slacikova M, Panait MLC, Hellenbrand W, de Melker H, Sobotova Z, Icardi G, Andrews N, Pebody R, Van Damme P, Kafatos G, Miller E, & Hatzakis A (2009). A comparison of hepatitis B seroepidemiology in ten European countries. *Epidemiology and Infection* **137**, 961-969.

Nardone A, de Ory F, Carton M, Cohen D, Van Damme P, Davidkin I, Rota MC, de Melker H, Mossong J, Stacikova M, Tischer A, Andrews N, Berbers G, Gabutti G, Gay N, Jones L, Jokinen S, Kafatos G, de Aragón MV, Schneider F, Smetana Z, Vargova B, Vranckx R, & Miller E (2007). The comparative sero-epidemiology of varicella zoster virus in 11 countries in the European region. *Vaccine* **25**, 7866-7872.

Nardone A & Miller E (2004). Serological surveillance of rubella in Europe: European Sero-Epidemiology Network (ESEN2). *Eurosurveillance* **9**, article=3.

Nardone A, Pebody RG, van den Hof S, Lévy-Bruhl D, Plesner AM, Rota MC, Tischer A, Andrews N, Berbers G, Crovari P, Edmunds WJ, Gabutti G, Saliou P, & Miller E (2003). Sero-epidemiology of mumps in western Europe. *Epidemiology and Infection* **131**, 691-701.

Nardone A, Tischer A, Andrews N, Backhouse J, Theeten H, Gatcheva N, Zarvou M, Kriz B, Pebody RG, Bartha K, O'Flanagan D, Cohen D, Duks A, Griskevicius A, Mossong J, Barbara C, Pistol A, Slacikova M, Prosenc K, Johansen K, & Miller E (2008). Comparison of rubella seroepidemiology in 17 countries: progress towards international disease control targets. *Bulletin of World Health Organization* **86**, 118-125.

Ndumbe PM & Lvenisky RJ (1985). Immunological cross-reactivity among three herpesviruses. *Journal of Immunological methods* **83**, 337-342.

Neal S & Efstratiou A (2007). DIPNET - Establishment of a dedicated surveillance network for diphtheria in Europe. *Eurosurveillance* **12**, article=5.

Nelson KE & Masters Williams C (2007). *Infectious Disease Epidemiology* Jones and Barlett Publishers, Sudbury Massachusetts.

Osborne K, Gay N, Hesketh L, Morgan-Capner P, & Miller E (2000). Ten years of serological surveillance in England and Wales: methods, results, implications and action. *International Journal of Epidemiology* **29**, 362-368.

Osborne K, Weinberg J, & Miller E (1997). The European sero-epidemiology network. *Eurosurveillance* **2**, article=3.

Parker RA, Erdman DD, & Anderson LJ (1990). Use of mixture models in determining laboratory criterion for identification of seropositive individuals: application to parvovirus B19 serology. *Journal of Virological Methods* **27**, 135-144.

Pebody RG, Andrews N, Brown D, Gopal R, de Melker H, Francois G, Gatcheva N, Hellenbrand W, Jokinen S, Klavs I, Kojouharova M, Kortbeek T, Kriz B, Prosenc K, Roubalova K, Teocharov P, Thierfelder W, Valle M, Van Damme P, & Vranckx R (2007). The seroepidemiology of herpes simplex virus type 1 and 2 in Europe. *Sexually Transmitted Infections* **80**, 185-191.

Pebody RG, Edmunds WJ, Conyn-van Spaendonck M, Olin P, Berbers G, Rebiere I, Lecoeur H, Crovari P, Davidkin I, Gabutti G, Gerike E, Giordano C, Hesketh L, Plesner AM, Raux M, Rota MC, Salmaso S, Tischer A, Valle M, & Miller E (2000). The seroepidemiology of rubella in western Europe. *Epidemiology and Infection* **125**, 347-357.

Pebody RG, Edmunds WJ, Conyn-van Spaendonck M, Olin P, Berbers G, Rebiere I, Lecoeur H, Crovari P, Davidkin I, Gabutti G, Gerike E, Giordano C, Hesketh L, Plesner AM, Raux M, Rota MC, Salmaso S, Tischer A, Valle M, & Miller E (2001). Erratum. *Epidemiology and Infection* **126**, 333.

Pinot de Moira A & Nardone A (2005). Varicella zoster virus vaccination policies and surveillance strategies in Europe. *Eurosurveillance* **10**, article=2.

Plikaytis BD & Carlone GM (2005). Statistical considerations for vaccine immunogenicity trials. Part 1: introduction and bioassay design and analysis. *Vaccine* **23**, 1596-1605.

Plotkin SA (2010). Correlates of protection induced by vaccination. *Clinical and Vaccine Immunology* **17**, 1055-1065.

ProImmune Limited. Protocols for the preparation of blood plasma and serum.

http://www.proimmune.com/ecommerce/pdf_files/PR31.pdf . 2009. Access date. 2-8-2010.

Ref Type: Electronic Citation

Qin L, Gilbert PB, Corey L, McElrath MJ, & Self SG (2007). A framework for assessing immunological correlates of protection in vaccine trials. *Journal of Infectious Diseases* **196**, 1304-1312.

Reynolds MA, Chaves SS, Harpaz R, Lopez AS, & Seward JF (2008). The impact of the varicella vaccination program on herpes zoster epidemiology in the United States: A review. *Journal of Infectious Diseases* **197**, 635-780.

Richardson DB & Ciampi A (2003). Effects of exposure measurement error when an exposure variable is constrained by a lower limit. *American Journal of Epidemiology* **157**, 355-363.

Roncalli T & Lagache T. A Gauss implementation of skew Normal/Student distributions (SN, ST, MSN and MST). The Skew library. <http://www.thierry-roncalli.com/download/gauss-skew.pdf> . 2004. Access date. 6-7-2011.

Ref Type: Electronic Citation

Saunders. Dorland's medical dictionary for health consumers. <http://medical-dictionary.thefreedictionary.com> . 2007. Accessed date. 2-8-2010.

Ref Type: Electronic Citation

Schafer JL. The multiple imputation FAQ page. <http://www.stat.psu.edu/~jls/mifaq.html> . 2010. Accessed date. 6-10-2010.

Ref Type: Electronic Citation

Schisterman EF, Vexler A, Whitcomb BW, & Liu A (2006). The limitations due to exposure detection limits for regression. *American Journal of Epidemiology* **163**, 374-383.

Schwarz NG, Bernard H, Melnic A, Bucov V, Caterinciuc N, An der Heiden M, Andrews N, Pebody R, Aidyralieva C, & Hahné S (2010). Mumps outbreak in the Republic of Moldova, 2007-2008. *Journal of Pediatric Infectious Diseases* **29**, 703-706.

Sheppard C, Cohen B, Andrews N, & Surridge H (2001). Development and evaluation of an antibody capture ELISA for detection of IgG to Epstein-Barr virus in oral fluid samples. *Journal of Virological Methods* **93**, 157-166.

Spika JS (2006). Measles elimination 2010 target: The need to meet the specific risk group. *Eurosurveillance* **11**, article=1.

Steffens I, Martin R, & Lopalco PL (2010). Spotlight on measles 2010: measles elimination in Europe - a new commitment to meet the goal by 2015. *Eurosurveillance* **15**, article=1.

Talukder Y, Gobal R, Andrews N, Glenn M, Breuer J, & Brown D (2005). Development and evaluation of varicella zoster virus ELISA for oral fluid suitable for epidemiological studies. *Journal of Virological Methods* **128**, 162-167.

Talukder Y, Kafatos G, Pinot de Moira A, Aquilina J, Parker SP, Crowcroft NS, Brown DW, & Breuer J (2007). The seroepidemiology of varicella zoster virus among pregnant Bangladeshi and white British women in the London Borough of Tower Hamlets, UK. *Epidemiology and Infection* **135**, 1344-1353.

Taylor GP, Bodéus M, Courtois F, Pauli G, Del Mistro A, Machuca A, Padua E, Andersson S, Goubau P, Chieco-Bianchi L, Soriano V, Coste J, Ades AE, & Weber JN (2005). The seroepidemiology of human T-lymphotropic viruses: types I and II in Europe: a prospective study of pregnant women. *Journal Acquired Immune Deficiency Syndrome* **38**, 104-109.

The Gale Group. Gale encyclopedia of medicine. <http://medical-dictionary.thefreedictionary.com> . 2008. Access date. 2-8-2010.

Ref Type: Electronic Citation

Tischer A, Andrews N, Kafatos G, Nardone A, Berbers G, Davidkin I, Aboudy Y, Backhouse J, Barbara C, Bartha K, Bruckova B, Duks A, Griskevicius A, Hesketh L, Johansen K, Jones L, Kuersteiner O, Lupulescu E, Mihneva Z, Mrazova M, de Ory F, Prosenc K, Schneider F, Tsakris A, Smelhausova M, Vranckx R, Zarvou M, & Miller E (2007). Standardisation of measles, mumps and rubella assays to enable comparisons of

seroprevalence data across 21 European countries and Australia. *Epidemiology and Infection* **135**, 787-797.

Tong DDM, Buxser S, & Vidmar TJ (2007). Application of a mixture model for determining the cutoff threshold for activity in high-throughout screening. *Computational Statistics and Data Analysis* **51**, 4002-4012.

Tookey P (2004). Rubella in England, Scotland and Wales. *Eurosurveillance* **9**, article=1.

van Hoek AJ, Gay N, Melegaro A, Opstelten W, & Edmunds WJ (2009). Estimating the cost-effectiveness of vaccination against herpes zoster in England and Wales. *Vaccine* **27**, 1454-1467.

Vandermeulen C, Leroux-Roels G, & Hoppenbrouwers K (2009). Mumps outbreaks in highly vaccinated populations. *Human Vaccines* **5**, 494-496.

Vilayleck MS (2002). The WHO European Region declared free of polio. *Eurosurveillance* **7**, article=3.

von Hunolstein C, Aggerbeck H, Andrews N, Berbers G, Fivet-Groyne F, Maple PAC, Olander RM, Raux M, & Tischer A (2000). European Sero-Epidemiology Network: standardisation of the results of diphtheria antitoxin assays. *Vaccine* **18**, 3287-3296.

Vyse A, Pebody R, & Hesketh L (2007). Interpreting the mumps epidemic in England in the early 21st century: a disease iceberg that is likely to continue? *International Journal of Antimicrobial Agents* **29**, S288.

Vyse AJ, Brown DW, Cohen BJ, Samuel R, & Nokes DJ (1999). Detection of rubella virus-specific immunoglobulin G in saliva by an amplification-based enzyme-linked immunosorbent assay using monoclonal antibody to fluorescein isothiocyanate. *Journal of Clinical Microbiology* **37**, 391-395.

Vyse AJ, Gay NJ, Hesketh LM, Morgan-Capner P, & Miller E (2004). Seroprevalence of antibody to varicella zoster virus in England and Wales in children and young adults. *Epidemiology and Infection* **132**, 1129-1134.

Vyse AJ, Gay NJ, Hesketh LM, Pebody R, Morgan-Capner P, & Miller E (2006). Interpreting serological surveys using mixture models: the seroepidemiology of measles, mumps and rubella in England and Wales at the beginning of the 21st century. *Epidemiology and Infection* **134**, 1303-1312.

Vyse AJ, Gay NJ, White JM, Ramsay ME, Brown DWG, Cohen BJ, Hesketh LM, Morgan-Capner P, & Miller E (2002). Evolution of surveillance of measles, mumps and rubella in England and Wales: Providing the platform for evidence-based vaccination policy. *Epidemiologic Reviews* **24**, 125-136.

Walsh CT & Fischbach MA (2009). New ways to squash superbugs. *Scientific American* **301**, 44-51.

Wearing HJ & Rohani P (2009). Estimating the duration of pertussis immunity using epidemiology signatures. *PLOS Pathogens* **5**.

Whitcomb BW & Schisterman EF (2008). Assays with lower detection limits: implications for epidemiological investigations. *Paediatric and Perinatal Epidemiology* **22**, 597-602.

WHO Regional Office for Europe. Operational targets for EPI diseases.

http://whqlibdoc.who.int/euro/1994-97/EUR_ICP_CMDS_01_01_14_Rev.1.pdf . 1996.

Access date. 1-9-2010.

Ref Type: Electronic Citation

WHO Regional Office for Europe. Surveillance guidelines for measles, rubella and congenital rubella syndrome in the WHO Region.

http://www.euro.who.int/_data/assets/pdf_file/0018/79020/E93035.pdf . 2009. Access date. 4-2-2011.

Ref Type: Electronic Citation

Wijsman RA (1996). New algorithms for the function $T(h,a)$ of Owen with application to bivariate normal and noncentral t -probabilities. *Computational Statistics and Data Analysis* **21**, 261-271.

Wilcox R (1998). Kolmogorov-Smirnov test. In *Encyclopedia of Biostatistics.*, eds.

Armitage P & Colton T, pp. 2174-2176. John Wiley & Sons Ltd, Chichester.

World Health Organization. The world health report – bridging the gaps.

http://www.who.int/whr/1995/en/whr95_ch1_en.pdf . 1995. Access date. 1-9-2010.

Ref Type: Electronic Citation

World Health Organization. Hepatitis A.

http://www.who.int/csr/disease/hepatitis/HepatitisA_whocdscsredc2000_7.pdf . 2000.

Access date. 8-9-2010.

Ref Type: Electronic Citation

World Health Organization (2005). Pertussis vaccines. *Weekly epidemiological record* **80**, 31-39.

World Health Organization (2006). Diphtheria vaccine. *Weekly epidemiological record* **3**, 24-32.

World Health Organization. Hepatitis B.

<http://www.who.int/mediacentre/factsheets/fs204/en/index.html> . 2008a. Access date. 8-9-2010a.

Ref Type: Electronic Citation

World Health Organization. Mumps - the disease.

<http://www.who.int/immunization/topics/mumps/en/index1.html> . 2008b. Access date. 8-9-2010b.

Ref Type: Electronic Citation

World Health Organization. Rubella - the disease.

<http://www.who.int/immunization/topics/rubella/en/index1.html> . 2008c. Access date. 8-9-2010c.

Ref Type: Electronic Citation

World Health Organization (2008d). *The global burden of disease: 2004 update*, pp. 54.

World Health Organization.

World Health Organization. Global immunization data.

http://www.who.int/immunization/newsroom/GID_english.pdf . 2009. World Health Organization. Access date. 5-10-2010.

Ref Type: Electronic Citation

World Health Organization Regional Office for Europe (2008). European immunization week 2008: progress towards regional goals. *Eurosurveillance* **13**, article=2.

Zar JH (1999). *Biostatistical Analysis*, 4th ed. Prentice-Hall, New Jersey, USA.

Zhang Z & Sun J (2010). Interval censoring. *Statistical Methods in Medical Research* **19**, 53-70.

Appendix I Programmes

Maximum likelihood in Stata

The maximum likelihood programmes (.ado files in Stata) were run using the Stata maximum likelihood estimation program `ml`.

(A) Programmes for censored data

The programmes shown below can be combined to provide more flexibility than the Stata built-in commands `cnreg` and `intreg`. For example, sigmoid censored regression models can be fitted for interval censored data.

The \log_{10} -transformed serological for the testing laboratory and the reference centre are included in variables `y` and `x`, respectively. Variable `cens` indicates whereas data are continuous (`cens = 0`), left-censored (`cens = -1`) or right-censored (`cens = 1`).

Maximum likelihood for a censored regression model (equivalent to cnreg command)

```
program cenlin                                     /* program cenlin.ado produces parameter
                                                    estimates for constant (alpha), slope
                                                    (beta) and standard deviation (sigma) for
                                                    a simple linear censored regression
                                                    model */

                                                    .

args lnf alpha beta sigma                        /* defining model parameters */

qui replace `lnf' = ln( normalden(y, `alpha' + `beta' * x, `sigma') ) if cens == 0
                                                    /* non-censored data */

qui replace `lnf' = ln( normal( (y - (`alpha' + `beta' * x)) / `sigma' ) ) if cens == -1
                                                    /* left-censored data */

qui replace `lnf' = ln( 1 - normal( (y - (`alpha' + `beta' * x)) / `sigma' ) ) if cens == 1
                                                    /* right-censored data */

end
```


Maximum likelihood for a sigmoid censored regression model

```
program censigmoid                                /* program censigmoid.ado produces
                                                    parameter estimates alpha, beta, gamma,
                                                    delta and standard deviation (sigma) for
                                                    a censored regression model of sigmoid
                                                    type */

args lnf alpha beta gamma delta sigma            /* defining model parameters */

qui replace `lnf' = ln( normalden(y, `alpha' + (`beta' / (1 + exp( - (`gamma' + `delta' * x) ) ) ), ///
`sigma') ) if cen == 0                          /* non-censored data */

qui replace `lnf' = ln( normal( y -(`alpha' + (`beta' / (1 + exp( - (`gamma' + `delta' * x) ) ) ) ) ///
`sigma') ) if cens == -1                        /* left-censored data */

qui replace `lnf' =ln( 1- normal(y - ( `alpha' + (`beta' / (1 + exp( - (`gamma' + `delta' * x) ) ) ) ) ///
`sigma') ) ) if cens == 1                      /* right-censored data */

end
```

Maximum likelihood for interval censored regression model (equivalent to intreg command)

```
program intlin                                /* program intlin.ado produces parameter
                                              estimates for constant (alpha), slope
                                              (beta) and standard deviation (sigma) for
                                              a simple linear interval censored
                                              regression model */

args lnf alpha beta sigma                    /* defining model parameters */

qui replace `lnf' = ln( normal( (high - (`alpha' + `beta' * x) ) / `sigma') ///
                      - normal( (low - (`alpha' + `beta' * x) ) / `sigma') ) if cen == 0
                                              /* interval censored data */

qui replace `lnf' = ln( normal( (inty - (`alpha' + `beta' * x) ) / `sigma') ) if cen == -1
                                              /* left-censored data */

qui replace `lnf' = ln(1 - normal( (inty - (`alpha' + `beta' * x) ) / `sigma') ) if cen == 1
                                              /* right-censored data */

end
```

(B) Programme for multiple imputations

The following code shows an example of how censored data can be imputed in Stata.

The log₁₀-transformed serological for the testing laboratory and the reference centre are included in variables y and x, respectively in the file "panel dataset.dta". Variable cen indicates whether data are continuous (cen = 0), left-censored (cen = -1) or right-censored (cen = 1).

* fitting a simple censored regression model like the one shown in Appendix I(A)

```
matrix initial=[-0.3\1.1\0.25]           /* initial values given for the maximum
                                          likelihood command (assuming a straight
                                          line */

ml model lf cenlin /alpha /beta /sigma    /* calling the maximum likelihood
                                          programme cenlin */

ml init initial, copy

ml maximize                               /* maximisation of the model */

mat cens=e(b)                             /* saving model estimates in a 1 x 3 matrix
                                          named cens where cens[1,1] is the
                                          constant, cens[1,2] is the slope and
                                          cens[1,3] the standard deviation */

loc m=10                                  /* defining the number of imputations */

mat mi=J(`m',3,0)                         /* creating a zero m x 3 matrix mi for
                                          saving the results of the 3 estimates and
                                          m imputations */
```

<pre> g fit=cens[1,1]+cens[1,2]*x </pre>	<pre> /* generating variable that includes the fitted values based on the censored regression estimates */ </pre>
<pre> local i=1 while `i'<=`m' { </pre>	<pre> /* beginning of m imputations */ </pre>
<pre> set seed `i' qui g simres`i'=(invnorm(uniform()))*cens[1,3] </pre>	<pre> /* generating results from a normal distribution with mean = 0 and standard deviation = cens[1,3] which as estimated from the censored regression model above */ </pre>
<pre> qui g simy`i'=fit+simres`i' </pre>	<pre> /* generating the simulated data */ </pre>
<pre> qui drop simres`i' </pre>	
<pre> qui replace simy`i'=y if cen==0 </pre>	<pre> /* for non-censored data replace the original data i.e. keep the imputed data only for censored observations */ </pre>
<pre> qui regress simy`i' x </pre>	<pre> /* fit a linear regression on the imputed dataset */ </pre>
<pre> mat a=e(b) </pre>	<pre> /* save the model output in a 1 x 2 matrix named a, where a[1,2] is the constant and a[1,1] is the slope */ </pre>
<pre> loc b=sqrt(e(rss)/e(df_r)) </pre>	<pre> /* saving the variance as a local variable named b */ </pre>

```

* saving the estimates of each imputation in matrix mi

mat mi[`i',1]=a[1,2]
mat mi[`i',2]=a[1,1]
mat mi[`i',3]=`b'

local i=`i'+1

}                                /* end of m imputations */

clear

svmat mi                        /* inserting the imputation estimates
                                saved in the matrix mi into a dataset of
                                m rows and 3 columns (mi1 for constant,
                                mi2 for slope and mi3 for standard
                                deviation) */

* calculation of the overall MI estimates
*

qui summ mi1
loc qmean1=r(mean)             /* saving constant estimate as the local
                                variable qmean1 */

qui summ mi2
loc qmean2=r(mean)             /* saving slope estimate as the local
                                variable qmean2 */

* calculation of the overall standard error (within-imputation)
*

qui summ mi3
loc umean=r(mean)              /* saving standard error estimate as the
                                local variable umean */

```

```

* calculation of the between-imputation variance
*

qui g bm1=(mi1-`qmean1')^2
qui summ bm1
loc bmean1=r(sum)/(`m'-1)                                /* saving variance estimate for the
                                                            constant as the local variable
                                                            bmean1 */

qui g bm2=(mi2-`qmean2')^2
qui summ bm2
loc bmean2=r(sum)/(`m'-1)                                /* saving variance estimate for the
                                                            slope as the local variable
                                                            bmean2 */

* calculating the total variances
*

loc sd1=sqrt(`umean'+(1+(1/`m'))*`bmean1')                /* saving the total variance estimate for
                                                            the constant as the local variable sd1 */

loc sd2=sqrt(`umean'+(1+(1/`m'))*`bmean2')                /* saving the total variance estimate for
                                                            the slope as the local variable sd2 */

* calculating and displaying the MI estimate together with the 95% CI for the constant
loc df1=(`m'-1)*((1+((`m'*`umean')/((`m'+1)*`bmean1'))))^2
di "alpha= " `qmean1'
di "95% CI= " `qmean1'-((invttail(`df1',0.025))*`sd1') ", " `qmean1'+((invttail(`df1',0.025))*`sd1')

* calculating and displaying the MI estimate together with the 95% CI for the slope
loc df2=(`m'-1)*((1+((`m'*`umean')/((`m'+1)*`bmean2'))))^2
di "alpha= " `qmean2'
di "95% CI= " `qmean2'-((invttail(`df2',0.025))*`sd2') ", " `qmean2'+((invttail(`df2',0.025))*`sd2')

```

(C) Programme for mixture models

The programme norm2age4 is an example of how the mixture model described in Equation 7.3 can be implemented in Stata.

The \log_{10} -transformed results of the serosurvey are included under variable y, whereas variable cens indicates whereas data are continuous (cens = 0), left-censored (cens = -1) or right-censored (cens = 1). Variable agegrp2 indicates which age group the individual sample belongs to.

```
program norm2age4                                /* file norm2age4.ado */

args lnf m1 s1 m2 s2 p1 p2 p3 p4                /* model parameters - lnf stands for log
                                                likelihood, m1 s2 is the mean and
                                                standard deviation of the negative
                                                component, respectively, and similarly
                                                m2, s2 the mean and standard deviation
                                                of the positive component. p1 - p4 are the
                                                proportion positives for 4 age groups */

loc i = 1

while `i' <= 4 {                                /* maximising over 4 age groups */

qui replace `lnf' = ln ( (1 - `p`i') * normalden (y, `m1', `s1') + `p`i' * normalden (y, `m2', `s2') ) ///
               if cens == 0 & agegrp2 == `i'      /* non-censored data */

* left-censored data

qui replace `lnf' = ln ( (1 - `p`i') * normal ( (y - `m1') / `s1') + `p`i' * normal ( (y - `m2') / `s2') ) ///
               if cens == -1 & agegrp2 == `i'     /* left-censored data */
```

```

* right-censored data

qui replace `lnf' = ln ( (1 - `p`i") * (1 - normal ( (y - `m1') / `s1') ) + ///
                    `p`i" * (1 - normal ( (y - `m2') / `s2') ) ) if cens==1 & agegrp2==`i'
                                /* right-censored data */

loc i = `i' + 1

}

end

```


(D) Programme for cut-off estimation using R (Equation 8.5)

The following code is an implementation in R of Equation 8.5 for estimating assay cut-off using the mixture model described in Section 8.5.

```
# The skew normal distribution library sn was used
```

```
library (sn)
```

```
# parameter estimates as given by the mixture model in Section 8.5
```

```
ma = -2.214302
```

```
sa = 0.5014619
```

```
mb1 = 0.6253432
```

```
mb2 = 0.5231688
```

```
mb3 = 0.4518782
```

```
mb4 = 0.3778452
```

```
sd1 = 0.47796613
```

```
sd2 = 0.52420586
```

```
sd3 = 0.48913248
```

```
sd4 = 0.49233756
```

```
l1 = -2.807696
```

```
l2 = -2.186895
```

```
l3 = -1.686308
```

```
l4 = -1.616152
```

```
p1 = 0.32560735
```

```
p2 = 0.75491795
```

```
p3 = 0.92710461
```

```
p4 = 0.94356571
```

```

# matrix for saving results
results <- matrix(0, nrow = 1, ncol = 10)

# age group 1
#
hc = function (x) {
# defining function hc that needs to be
# solved (Equation 8.4) for age group 1

fneg = pnorm ( (x - ma) / sa) # normal density for negative component

fpos = psn (x, location = mb1, scale = sd1, shape = l1)
# skew normal density for positive
# component

y = abs ( (p1 * fpos) - ( (1 - p1) * (1 - fneg) ) ) # Equation 8.4
return (y)
}

co = optimize(hc, lower = -2, upper = 0) # function optimize finds the local
# minimum or maximum within the area
# specified (in this case (-2, 0))

co1 = co$minimum # local minimum of age group 1 stored in
# co1

```

```

# age group 2
#
hc = function(x) {                                     # defining function hc that needs to be
                                                         # solved (Equation 8.4) for age group 2

    fneg = pnorm ( (x - ma) / sa)

    fpos = psn (x, location = mb2, scale = sd2, shape = l2)

    y = abs( (p2 * fpos) - ( (1 - p2) * (1 - fneg) ) )

    return (y)

}

co = optimize(hc, lower = -2, upper = 0)

co2 = co$minimum                                       # local minimum of age group 2 stored in
                                                         # co2

```

```

# age group 3
#
hc = function (x) {                                     # defining function hc that needs to be
                                                         # solved (Equation 8.4) for age group 3

    fneg = pnorm ( (x - ma) / sa)

    fpos = psn (x, location = mb3, scale = sd3, shape = l3)

    y = abs( (p3 * fpos) - ( (1 - p3) * (1 - fneg) ) )

    return (y)

}

co = optimize(hc, lower = -2, upper = 0)

co3 = co$minimum                                       # local minimum of age group 3 stored in
                                                         # co3

```

```

# age group 4

#
hc = function (x) {                                     # defining function hc that needs to be
                                                         # solved (Equation 8.4) for age group 4

    fneg = pnorm ( (x - ma) / sa)

    fpos = psn (x, location = mb4, scale = sd4, shape = l4)

    y = abs( (p4 * fpos) - ( (1 - p4) * (1 - fneg) ) )

    return (y)

}

co = optimize(hc, lower = -2, upper = 0)

co4 = co$minimum                                         # local minimum of age group 4 stored in
                                                         # co4

# displaying the cut-off estimates for the four age groups

co1

co2

co3

co4

```

PAGES NOT SCANNED AT THE
REQUEST OF THE UNIVERSITY

SEE ORIGINAL COPY OF THE THESIS
FOR THIS MATERIAL